

Bashing Japan/Bashing America

President George Bush's mission to Japan last week began as if it were the most serious trade mission ever mounted from the United States, but it proved predictably a flop.

WHEN US President George Bush left Washington for an Asian tour ending at Tokyo, he said his mission was to expand free trade in the Pacific rim in order to create "jobs, jobs, jobs" for US workers—particularly in the automobile industry. In fact, the White House went so far as to blame Japan for the United States current economic recession. But by the time the trip was over, Bush had merely succeeded in making the US automobile industry look to all the world like the dying giant it is. Bashing Japan did not work.

After days of what were apparently tense negotiations with Japanese prime minister Kiichi Miyazawa, Bush (the herald of free-market trading) won a Japanese promise to buy an additional 20,000 US-built automobiles in 1994. Not only does such an agreement smack of a managed market, but it also raises an embarrassing question: why would Japanese consumers want to buy US cars when their own are better and less expensive? True, some Japanese buy luxury cars from Detroit as a matter of status, but not in numbers sufficient to affect trade policy, and certainly not sufficient to create jobs in the US automobile industry.

In short, the emphasis Bush placed on selling cars to Japan was a mistake (as was his decision to travel in the conspicuous company of the chairmen of the three US automobile companies) and the results of his negotiations are a joke. The chairman of General Motors (GM), who is closing 21 plants and laying off 74,000 workers, wants to sell more GM cars in Japan, but seems not yet to have figured out that they need to be built with the steering wheel on the right-hand side. He said GM might reconsider if its market share were high, but he is putting the cart before the horse. Then, upon returning home, Chrysler chairman Lee Iacocca had the audacity to say, "We do not have idiots running General Motors, Ford and Chrysler". On the record of this trip, who will believe him?

Bush's major error in planning his trip to Japan lay in failing to select leaders of the best in US industry as his unofficial ambassadors. Where were the heads of the pharmaceutical companies that do millions of dollars in trade with Japan on the strength of their high-quality products? Where were the scientific and business leaders of the biotechnology industry, who represent a vital link to the future of international trade? Where were the people from the US computer industry, which is still at the intellectual forefront, even if Japan has succeeded brilliantly

at capturing the consumer market?

Perhaps one of the brightest (but least acknowledged) accomplishments of Bush's trip was a computer trade agreement that will give US companies the opportunity to compete in Japan for the sale of personal and minicomputers, as well as mainframes. Industry experts estimate that the agreement could be worth as much as \$2,000 million in sales and is far more important than a meaningless plan to foist 20,000 cars on Japanese consumers who may well refuse to buy them. Consumers in Tokyo already have expressed resentment at Detroit's presumption they will buy American cars.

For all its supposed weaknesses, the US position in science and technology research remains strong and should have been the centrepiece of the Bush trip for economic as well as diplomatic reasons. Instead, Bush chose to lay his prestige on the line for Detroit, which does not deserve it. Bush has described his dramatic fainting spell and illness at Miyazawa's banquet as an embarrassment, but there is no embarrassment in getting the flu. The embarrassment is in failing to put the right science and technology issues on his agenda. Next time, Bush would do well to consult his science adviser, his competitiveness council and even his trade representative — Mrs Carla Hills was conspicuous by her absence from Tokyo — if he wants to avoid embarrassing himself, not to mention the United States. □

The planet vanishes

Last summer's planet around a neutron star seems to be a computational artefact.

THE letter on page 213 from Andrew Lyne and M. Bailes at Jodrell Bank will occasion great disappointment to all kinds of people. Last July, the same group described what seemed to be the first incontrovertible evidence of a planet associated with a star other than the Sun. To be sure, the circumstances were unusual: the planet was in orbit about a neutron star, recognized by its rapid pulsation in the millisecond range. And the evidence, while unambiguous, was circumstantial; there is no sense in which the planet could have been 'seen'. Its existence was, rather, inferred from the periodic variations of the times at which radio pulses were received at Jodrell Bank. At different times over an interval of three years, the time at which radio



pulses reached the Earth differed from that expected by up to 10 milliseconds, implying a maximum displacement of about 3,000 km in the position of the neutron star from the centre of gravity of the planetary system.

Now, Lyne and Bailes say that their inference was mistaken. The measurements are correct, they say, but the periodic fluctuations in the times at which pulses are received do not arise from the periodic displacement of the pulsar by an orbiting planet. Instead, they are a consequence of the departure of the Earth's orbit about the Sun from strict circularity. That should have been allowed for in the computer programme used to reduce the data, but was not. The embarrassment with which Lyne and his colleagues must contend is all the greater because of the wide attention their inference has naturally received.

It is to be hoped that they will not be too downcast by this turn of events. On the contrary, they should take pride in the directness of their acknowledgement this week of their mistake. In no sense is their explanation obfuscated by irrelevancies, or by technical language from which the nature of the underlying error would have been hidden from all except those working in the field. No doubt Lyne, who has the unenviable task of explaining what went wrong to the American Astronomical Society at its meeting this week at Atlanta, will be similarly forthright on that occasion. His audience should recognize that his forthright retraction of the original claim is a model of how these things should be done.

Lyne might also legitimately argue that, even if the original claim was mistaken, the idea that planets might be found by searching for departures from regularity of the arrival times of radio pulsations has already been a stimulus to others working in the field. The report last week (*Nature* 355, 145; 9 January 1992) by Wolszczan and Frail, that yet another millisecond pulsar is accompanied by at least two planets, is based on the same technique. Those concerned will now, no doubt, be poring over the details of their computer reduction in case they also incorporate invalid assumptions; if there are mistakes, they must be different from those responsible for the six-month orbital period of the planet now disappeared. If the reductions prove to be robust, this will not be the first occasion on which a false result has stimulated others to find a valid replication of it. The world at large should understand that such ups and downs are inseparable from science. □

A zoo with animals

The Zoological Society of London seems still uncertain about its running of the London Zoo.

THERE is no requirement in natural law that every metropolitan city should have a public zoo, but many do. London, in its time the centre of nineteenth-century speculation about the faunal wonders of the natural world, has had one since 1828, thanks to the Zoological Society

of London, a venerable scientific society which, over the decades, has stumbled into being a theme park operator as well. For at least 20 years, the tail (represented by the zoo) has firmly wagged the dog (of the Zoological Society). City-centre zoos have proved to be expensive to operate. Government subsidies have been sought (and sometimes won). But, last year, the professional management recruited to run the London Zoo began to circulate the rumour that its site in Regent's Park would be evacuated of permanent animal residents, and turned into a theme park devoted to conservation causes. Predictably, there were ructions. On Monday last week, an extraordinary general meeting of the society seems to have derailed that plan. What will happen now?

The council of the Zoological Society seems to have determined on a strategy intermediate between a theme park of Disney proportions and the abandonment of the site at Regent's Park. Some animals will remain on the site, there will be exhibitions analogous with those found in modern museums to illuminate issues in biology and there will be a more vigorous breeding programme aimed at the conservation of endangered species. All that makes sense. Among other things, it will let the London Zoo remain the medium-sized business the organization can handle without excessive pomp, will blunt the edge of the criticism of those who hold that animals should never be kept in zoos and will also help to ensure the continued place of the London Zoo in the international scheme of things.

But that is not enough. Paradoxically, the London Zoo has managed to get into a money muddle just when public interest in animal species and their conservation is greater than ever. Is there no way in which that public interest can be tapped for the benefit of a permanent exhibition at Regent's Park? Precedents elsewhere, notably at San Diego, suggest that there is plenty that could be done. What the London Zoo needs is a scheme allowing people who are innocent of qualifications in formal zoology to become full members of whatever science-based organization eventually runs the zoo. For that purpose, the Zoological Society should either admit and encourage non-zoologists to equal membership (there are already two categories of membership) or found a new organization ('The London Zoo Society'? 'Friends of Fauna'?) to which ownership and management of the zoo could be transferred.

What there is now is simply a scheme allowing private persons to contribute to the zoo financially at various rates in return for free admission on specified occasions and, at the higher rates of contribution, dining rights in the zoo's well-appointed restaurants. That represents an effete British way of tackling an urgent problem. Those who seek to eat lunch at the zoo are many fewer than those with a serious and enthusiastic interest in the world's fauna, and may not even be a subset thereof. The radicals at last week's meeting believe that they won "the war, if not the battle", but it is not the war that really matters. □

Battle lines form over fetal tissue research

- New research coalition mounts lobby offensive
- Two-thirds of Congress is distant goal

Washington

THIRTY-SEVEN research and health groups, joined under a banner of "research freedom", are planning to mount the strongest attack yet on a four-year-old US moratorium on federally funded research on fetal tissue for human transplantation.

Spurred by support in the House of Representatives, where a bill that would overturn the moratorium passed by a large margin last summer, the members of the new Coalition for Research Freedom are rallying behind similar legislation in the Senate. Known as the Research Freedom Act of 1991, and introduced by Senator Brock Adams (Democrat, Washington), the Senate bill is due to be incorporated into legislation reauthorizing the National Institutes of Health (NIH). The Senate Labor and Human Resources Committee is expected to vote on it early next month.

It is on the Senate floor, however, that the research advocates will face their most difficult hurdle. The support of a majority of the lawmakers will probably not be enough to enact the legislation; a two-thirds advantage is needed to override a threatened presidential veto. Last July, the House bill—introduced by Representative Henry Waxman (Democrat, California)—passed 274–144: a clear majority, but not two-thirds. To get the required votes in the Senate, supporters of the Adams bill will have to break partisan lines and convince traditionally anti-abortion legislators that the fetal tissue research question is not simply the abortion question in a scientific disguise.

One example of the kind of convert that the research groups need is Mark Hatfield, an anti-abortion Republican senator from Oregon. In a letter dated 26 November to the American Federation for Clinical Research, one of the members of the Washington-based coalition, Hatfield explained why he had decided to break with the anti-abortion lobby and back the fetal-tissue measure: "Given the great promise of fetal tissue transplants . . . I believe the truly 'pro-life' position on this issue is that of supporting the research."

Coalition members are now conducting a 'straw poll' to test the political winds in Congress. But there is a difference between philosophical support for a worthy cause, and actually being the swing vote responsible for President George Bush's first veto defeat. So far, the coalition's attempts to separate fetal-tissue research from the abortion debate have been partially successful; supporters of the

Waxman bill included some 70 'pro-life' congressmen. Although they are encouraged by their success, coalition members harbour no illusions about the difficulty in getting votes to sustain a veto.

"It's been a slow process of education," says Joan Samuelson, a California lawyer with Parkinson's disease who has been one of the strongest supporters of



Bernadine Healy supports fetal research, but as NIH director must support the moratorium.

legislation overturning the moratorium on fetal tissue transplants, which are being explored as a possible cure for Parkinson's. Rather than focusing on the ethics of fetal tissue research, the coalition is emphasizing its scientific benefits and the safeguards in the legislation that they say would deter women from having abortions in order to donate fetal tissue to science. Opponents of fetal tissue transplantation argue, against available evidence, that the research would increase the rate of elective abortions.

Beginning with the report of a 1988 advisory committee to the National Institutes of Health (NIH), proponents of fetal tissue research have devised elaborate procedures to ensure that a woman would not know to whom tissue from her fetus would go, nor receive any payment for the tissue. Such 'walls' between the research and the abortion decision are in the current legislation, as well. But anti-abortion activist argue that it is not enough not to

know the name of the potential recipient; just knowing that some good may come of an abortion could be enough to persuade an undecided woman, they say.

Because fetal tissue grows rapidly and is less likely that tissue from adults to be rejected by the body's immune system, it has been a focus of transplant research since the 1950s. But with the exception of its use in DiGeorge's syndrome, a rare digestive disorder, its effectiveness is still not clear. Research in Mexico, Sweden and the United States has shown some promising results in the treatment of Parkinson's disease, but the procedure is still experimental.

The US moratorium on fetal tissue research applies only to federally funded work on transplanting tissue from human fetuses. NIH director Bernadine Healy, who served on the NIH fetal tissue advisory panel before joining the agency, has testified that the agency is spending about \$8 million a year on fetal tissue research that is not covered by the moratorium, mostly for nontransplant research and animal tissue. A small amount of independent research is also funded by private sources, although, as Samuelson says, "the moratorium has placed a stigma on the work, making it look inappropriate" and discouraging further private support. Coalition members argue that permitting federally funded—and regulated—fetal tissue research will ensure that private efforts comply with the safeguards outlined in the congressional legislation.

Other research subjects that have fallen afoul of politics have a similar abortion link. *In-vitro* fertilization research is the subject of another Administration moratorium, on the grounds that several ova are usually fertilized in the process, although they are not all implanted in the womb. These discarded cells are tantamount to abortions, pro-life activists claim. And although RU-486, the abortion drug, may have medical application for breast cancer and Cushing's syndrome, Roussel-Uclaf, the French pharmaceutical company that produces it, has restricted its use in the US because of anti-abortion fervor.

Although the drive for support from two-thirds of the Congress is the current focus for lobbying, a lawsuit remains a possibility. Because the moratorium was introduced without the normal notice and comment period required for federal regulations, it may be in violation of the Administrative Procedures Act, which dictates how regulations must be made. Several groups, led by the Chicago-based United Parkinson Foundation, have threatened to sue the government. But before embarking on what promises to be a long and protracted legal battle, the coalition will concentrate on Congress in its effort to see the fetal tissue transplantation moratorium come to an end.

Christopher Anderson

Britain gives the green light

London

THE British government's Department of Health today (16 January) releases a long-awaited report expected to conclude that there are no serious ethical objections to somatic gene therapy. The report, from an independent committee chaired by the lawyer Sir Cecil Clothier, gives the green light to British research groups planning gene therapy experiments, and adds to the growing consensus in Europe that the ethical questions raised by the genetic manipulation of somatic cells are no more difficult than those surrounding organ transplantation.

Clothier's committee, set up two years ago, is thought have to made a clear distinction between somatic gene therapy and the manipulation of germ line cells. Like the other European ethical committees addressing the issue, led by the Dutch National Health Council which reported in late 1989, the British group is expected to recommend a moratorium on germ line therapy, which most geneticists say could lead to unpredictable effects in subsequent generations.

The Clothier report also comes just as innovation in gene therapy research, which began in the United States, expands to include Europe. Dinko Valerio is the Dutch

researcher who first cloned the adenoside deaminase (ADA) gene used in the pioneering gene therapy experiments of French Anderson and Michael Blaese, from the National Institutes of Health in Bethesda, Maryland (see *Nature* 346, 402; 1990). Valerio hopes later this year to begin a clinical trial that may herald the treatment of genetic disorders by transplanting manipulated bone marrow cells. So far, the NIH experiments with ADA deficiency — an extremely rare inherited disorder — have involved the transformation of relatively short-lived lymphocytes, meaning that the successful treatment would require cells to be repeatedly transfused on a regular basis.

Valerio, who works at the Netherlands Organization for Applied Research's applied radiobiology and immunology institute at Delft, has completed successful trials of his technique in mice and rhesus monkeys. In both cases, the ADA gene was inserted into cultured bone-marrow cells by using a viral vector. After transplanting the marrow cells into his experimental animals, Valerio found that the ADA gene was expressed in a range of blood cell types. He is now waiting for approval of an ethics committee of the Dutch National Health Council, before

going ahead with his human clinical trial.

In Britain, the Clothier report is unlikely to "open the floodgates" to a rush of gene therapy experiments, according to Sir David Weatherall, from the University of Oxford, who is a member of the Clothier committee; he says that few British research groups are ready to begin gene therapy. But Bob Williamson, whose cystic fibrosis research group at St Mary's Hospital in London is expected to be among the first to start, believes the report is important because it will inform and reassure the public. By distinguishing between somatic and germ line therapy, the report will point to "a middle course that will encourage scientists to go ahead sensibly," he says.

The expected activity of British geneticists in gene therapy research may introduce some new experimental approaches. In most experiments, genes have been inserted into somatic cells using viral vectors, but Williamson is exploring the use of artificial human chromosomes as vectors. The cystic fibrosis gene has been expressed in rats after transfer using an adenovirus vector, but there are some outstanding concerns about the safety of this delivery system arising from the possibility that viral vector DNA could recombine with wild type adenovirus sequences already present in the lungs.

Peter Aldhous

SPACE SCIENCE

Mute Magellan finds its voice

Washington

MAGELLAN, the radar mapping probe now orbiting Venus, failed for seven days last week before mission engineers cobbled together a repair that will allow the spacecraft to continue at 40 per cent capacity. The failure, which rendered the Magellan's main transmitter inoperable, came at the end of the orbiter's successful first coverage of Venus, but before it could start a second set of passes that would yield a three-dimensional view of the planet's surface.

Engineers at the project's National Aeronautics and Space Administration (NASA) headquarters first detected the failure on 4 January, when they found that the orbiter was returning a clean signal, but no data. At first concerned that something had gone wrong with the radar instruments, the engineers were relieved to find that the data immediately reappeared at the usual 268,000 bits per second when they switched to the backup transmitter. But their reprieve was not to last. As the on-board electronics for the backup warmed up, that transmitter developed a noisy whistle that obscured the data. NASA project officials believe that there is a crack in a component that enlarges as the part heats up.

During the next week, NASA engineers tried several different transmission frequencies, trying to separate the data from the noise. But each time, either the whistle, or one of its harmonics, landed squarely on the data. Finally, the engineers tried the transmitter at its lower transmission rate of 1,200 bits per second, normally used only for telemetry data. At

Martin Marietta Corp./Science Photo Library

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Magellan in trouble.

this rate (which uses a far lower carrier and data frequency) the data finally came through cleanly. Although the lower transmission rate is only 40 per cent of normal, and the main antenna appears to be out for good, the probe can at least communicate again.

Magellan, initially scheduled to map only 70 per cent of Venus's surface, had, until the glitch last week, surpassed all expectations. It has mapped more than 95 per cent of the planet, and was filling in the last five per cent when the transmitter failed. This week, project engineers will send the probe new computer instructions to avoid further damage.

On 17 January, the orbiter will take a week off to recycle its batteries, and then on 23 January, it will begin the third part of its mission, filling in remaining gaps in the surface mapping, viewing previously mapped areas from a new angle to provide a 'stereo' three dimensional effect, and double checking some inexplicable altitude measurements, including one of a cliff that may be the highest in the solar system and a possible sign of active seismic activity on the planet. But unless the NASA engineers are able to fix the main transmitter, something that now seems unlikely, Magellan will be able to transmit only about half of what it sees, which will force project scientists to pick their targets carefully.

Christopher Anderson

Challenge to British forensic database

London

THE British civil rights group Liberty is seeking to take London's Metropolitan Police to the European Court of Human Rights, questioning the legality of its database of DNA fingerprinting results. Liberty's argument turns on the case of Roy Williams, a man questioned during a 1988 murder inquiry and subsequently cleared after voluntarily submitting a DNA sample for analysis — but who later found that his DNA profile had been included in the Metropolitan Police database without his consent. If taken up by the European Commission for Human Rights, the case may take as long as five years to resolve, but Liberty's action highlights a debate likely to gain momentum over the coming

year, as national DNA profile databases come into general use in Britain and the United States. In Britain, the Home Office Forensic Science Service has set up a central database which police forces can ask to be trawled for matches to their scene-of-crime samples, and the Federal Bureau of Investigation (FBI) expects to have a similar US database in place a year from now.

John Wadham, Liberty's legal officer, complains that the police and forensic scientists in Britain are launching their DNA profile databases without waiting for a full public debate over what data should be included. He believes the establishment of databases should be governed by new legislation, pointing out

that, under the Police and Criminal Evidence Act, conventional fingerprint evidence must not be kept on file if a suspect is not convicted.

According to the Home Office (Britain's equivalent of an interior ministry), the Forensic Science Service computer database contains DNA profiles from convicted criminals and those collected during investigations that have not yet come to trial or which relate to unsolved crimes. After a trial, only the DNA profiles from those suspects who are convicted will remain in the computer — although a spokeswoman says that paper files may be retained. (British law provides far greater controls over the confidentiality of computer-held information than over paper records, but Wadham says the distinction is "irrelevant".)

Wadham says that the separate Metropolitan Police database seems to have been set up without even the rudimentary controls exercised by the Forensic Science Service: when he visited the force's Scotland Yard headquarters to ensure that Williams' DNA test results were deleted, Wadham says that the police found it necessary to call in a computer specialist; the system did not seem to have been set up to allow for the deletion of test results collected from innocent suspects. For their part, the Metropolitan Police refuse even to confirm that they possess a computerized DNA profile database.

John Hicks, head of the FBI's crime laboratory in Washington, is less worried about legal challenges to his agency's planned database. Only two classes of information will be included, he says: the DNA profiles of convicted felons, and crime scene evidence from cases that are still under investigation. Hicks adds that, by the time the FBI database comes into general use, it may have legislative backing. Representative Don Edwards' (Democrat, California) wide-ranging bill on forensic DNA testing (see *Nature* 351, 684; 1991), which would authorize the FBI to establish a DNA profile database, has now been passed by the House of Representatives.

Although the bill's progress is temporarily blocked by controversy surrounding other crime-related legislation to which it has become attached, Hicks believes it will be soon separated from the other measures, and passed into law. The situation in the United States could become more complicated if state and city authorities decide — like the Metropolitan Police in Britain — to set up their own DNA profile databases, but Hicks says he knows of no such plans.

Peter Aldhous

ANIMAL RESEARCH

Court favours mice, rats, birds

Washington

RULING that any federal regulation that fails to include rats, mice and birds as laboratory animals is "arbitrary and capricious", a federal court last week directed the US Department of Agriculture (USDA) to reconsider its interpretation of the Animal Welfare Act. The decision is a victory for animal-welfare activists, who had argued that USDA had overstepped its authority in choosing to exclude rodents and birds when writing regulations for implementation of the act (see *Nature* 351, 338; 1991).

Although the animal welfare act is silent on the subject of rats, mice and birds, it explicitly lists dogs, cats, monkeys, guinea pigs, hamsters, rabbits "live or dead" for inclusion in its guidelines and also calls for the inclusion of "such other warm-blooded animal" as the agriculture secretary finds is being used in research. The Secretary would not have to look far to find laboratory rats and mice in abundance.

In its defence, USDA had claimed that including the animals in its regulations would double or triple its inspection workload and force the agency to spend several million dollars to increase its staff. Officials have often argued that there is little evidence that laboratory rodents and birds — which, combined, make up more than 80 per cent of all laboratory animals — are being generally mistreated or kept in substandard housing. They point out that the animals are already covered by the guidelines of the Public Health Service (PHS), which funds most of the academic animal laboratories in the United States. Because rodents are so often used in research and so easily maintained, standard and humane caging is the norm, USDA says.

Animal welfare groups, including the Humane Society and the Animal Legal Defense Fund, which were both parties in the lawsuit, do not think that abuse of rats and mice is widespread. But they are concerned that no one is inspecting the facilities to watch for the exceptions. Because the PHS guidelines are not formal regulations like the Animal Welfare Act, the health service depends on institutions to police themselves.

In the decision, federal judge Charles Richey wrote that he "recognizes that enforcement of these regulations would require some expenditure of agency resources. Yet...the inclusion of rats, mice and birds under the Act would send an important message to those responsible for their care — that the care of these animals is something for which they are legally responsible...". He noted that USDA has not asked Congress for more money to expand its inspection programme; in fact, USDA once tried to have it eliminated.

If, as expected, USDA does modify its rules to include rats, mice and birds, most laboratories will have little trouble complying, says Barbara Rich, vice president of the National Association for Biomedical Research, an animal research lobby group. Rather than devise new standards for animal enclosures, USDA will probably just adopt the PHS guidelines as law, something it did with primates, dogs, cats and other research animals. Extending those rules to rats and mice "would mean some different record-keeping and reporting [procedures], but I don't see it as a big problem," Rich says. USDA has not decided if it will appeal the decision.

Christopher Anderson

Combating counterfeit drugs

TWENTY West African countries are launching a joint programme to combat the global trade in counterfeit medicines, believed now to be big business as well as a serious cause of needless death.

The campaign is backed by the United Nations' World Health Organization (WHO) in Geneva and its collaborating networks of research and training institutions worldwide. Pill piracy involves copies of medicines such as insulin, antibiotics and antivirals.

In one recent case, more than 100 Nigerian children aged under six years died after being given paracetamol, a pain killer, containing an industrial solvent.

Nigerian children suffering from malaria have also died after being treated with fake antimalaria drugs. Health authorities in Mexico have seized thousands of counterfeit burn remedies containing sawdust. In India, 14 hospital patients died recently after being treated with industrial rather than hospital grade glycerine. Hospitals and pharmacies in Europe have also been extensively abused by the forgers.

The health services of the poor countries, lacking training and testing facilities, are particularly vulnerable.

The West African initiative follows an agreement reached last year by representatives of 20 national health authorities at a regional meeting in Lomé, the capital of Togo, to combine resources against piracy. The meeting, brought together by WHO and the International Federation of Pharmaceutical Manufacturers Association, called for investment in education and quality control facilities. The programme will pay particular attention to the chronic shortages of medicines for tractable diseases, where forgers are likely to do most damage.

The programme builds on a pioneering experiment in Nigeria, financed by a \$68 million loan from the World Bank (equal to the country's annual health budget) when health workers were trained to identify fake or substandard medicines and laboratories established to carry out regular spot checks.

David Radel, a senior population and health specialist with the World Bank's Africa regional programmes, says that basic drugs such as antibiotics, antimalarials and agents to eliminate parasites are almost nonexistent, and that many of the drugs that can be had "are either fake, diluted or adulterated".

In the past decade, the illicit trade has become global, provoking judicial investigations in Europe and North America.

Much of the trade originates from developing countries that do not recognize the patents owned by the multinational

drug companies. But Thailand, whose substandard medical products have recently gained a global market, is about to change its policies on intellectual property, and other countries may well follow suit.

According to unofficial estimates by government health officials and drug company representatives, fraudulent products are creating substantial problems in many countries, including Bangladesh, Colombia, India, Indonesia, Brazil and Pakistan as well as Nigeria and its neighbours.

"All it takes is a person with access to a small laboratory, a total disregard for human beings and a touch of larceny in the heart," according to Milton Silverman, Mia Lydecker and Philip Lee of the Institute of Health Policy Studies, School of Medicine, University of California in San Francisco.

Some people estimate that fake drugs comprise at least a fifth of the products sold in Brazil's nonhospital pharmacies. Many state pharmacy inspectors are understood to supplement their \$115 monthly salary by helping druggists to ignore professional requirements.

In Indonesia, up to a third of the most widely used drugs may be fakes. Although the counterfeiters are easily caught, the usual punishment of three or four months imprisonment is hardly a deterrent, given the profits that can be earned.

The legitimate pharmaceutical companies, with sales of \$150,000 million a year, have hitherto been reluctant to acknowledge the problem for fear of losing consumer faith in their products. But some of them are beginning to cooperate with WHO to suppress the trade. The Swiss company Ciba-Geigy has provided specialist training, equipment and managerial advice in the establishment of a pharmaceutical quality control laboratory just opened in Yaoundé, the capital of Cameroon.

Nigeria's federal government has so far spent \$2 million on training and technology and set up specialist medical task forces for each of its 21 states. International assistance and the West African regional endeavour may well widen the programme and establish a development pattern helping other poor countries to control pill piracy.

Nigeria is using the World Bank loan to improve the supply of quality medicines in the states of Bendel, Cross River, Gongola and Kwara. The money will be invested in specialist training and laboratory facilities to revitalize the inadequate drug supply system and improve the way medicines are imported, manufactured, stored and distributed.

Thomas Land

Indian centre receives funds

New Delhi

A BOMBAY-BASED pharmaceutical company will fund biotechnology research at the International Centre for Genetic Engineering and Biotechnology (ICGEB) in New Delhi in what may prove to be an important contractual arrangement between an Indian private company and an international research centre.

Under an agreement reached with the United Nations Industrial Development Organization (UNIDO), Wockhardt Limited of Bombay will pay Rs 50 million (£1.2 million) to the biotechnology centre for development of five products which Wockhardt will have the right to market in India.

ICGEB director Dr K.K. Tiwari said that the Indian government has approved the agreement and the company has paid the first instalment of Rs 5 million to UNIDO. The rest will be paid over a period of five years.

Under the agreement, ICGEB will develop hepatitis-B vaccine, insulin, erythropoietin and two other biotechnology-based products. The Indian company will carry out toxicological studies and clinical trials before marketing the products commercially. Tiwari said work on those products has already begun and "we expect to start clinical trial on at least one product within a year".

ICGEB with two laboratories — one in New Delhi and the other at Trieste in Italy — was created by UNIDO to provide a training base for scientists in developing countries and to develop relevant products. Forty-four countries are members of ICGEB, but none except the host countries — Italy and India — have made any contributions for running the centre. Contracting research to a private company is viewed by UNIDO as a step to augmenting the resources and keeping ICGEB going until member countries start paying their dues.

The centre is three years old and, according to a spokesman for ICGEB, member states have two more years to decide whether or not to contribute. "For the present, our centre is not doing badly," the spokesman said. Besides the annual contribution of \$500,000 from India and \$1.2 million from Italy, ICGEB has received a \$400,000 grant from the Rockefeller Foundation for research in rice biotechnology. These, plus the grant of nearly \$2 million from Wockhardt, have brightened the chances of ICGEB's survival. Nevertheless, even the most committed optimist would be hard pressed to predict that ICGEB will ever receive funding from most of the member states.

K.S. Jayaraman

Japan vacillates over SSC

Tokyo

JAPANESE Prime Minister Kiichi Miyazawa has again demonstrated Japan's unwillingness to respond decisively to US requests for support of the Superconducting Super Collider (SSC).

Last week, in a summit meeting with President George Bush in Tokyo, Miyazawa agreed only to establish a joint US-Japan working group to study how Japan might participate in the SSC. This move, which postpones any decision on financial support until at least the end of this year, came despite widespread expectation in the past few weeks that Miyazawa would promise to contribute to the SSC (see *Nature* 355, 8; 2 January 1992). A few months ago, Japan's science-related ministries and agencies made it clear they cannot support the SSC out of their limited budgets, but the possibility of Miyazawa committing Japan to the project was left open.

Japan's vacillation over the SSC draws attention once again to the weakness of Japan's political leadership. The recent past is strewn with similar cases. For

example, both Miyazawa and his predecessor Toshiki Kaifu have made repeated attempts to set up a peace-keeping force in Japan to participate in United Nations activities, but to no avail. Miyazawa then tried unsuccessfully to introduce new taxes to fund international activities, such as the SSC. Even Yasuhiro Nakasone, who was considered one of Japan's strongest prime ministers in the postwar period, took three years to launch his pet project, the Human Frontier Science Program, and even then it was only a fraction of the size originally intended.

In the case of the SSC, the bureaucrats in the Japanese government are in large part opposed to contributing, and there is little support in the academic community,

but Japan was under strong political pressure to make Bush's visit to Tokyo a success. And Miyazawa apparently wanted to contribute to the SSC to ease pressure on other trade issues such as automobiles and rice. But in the final hour, he did not.

The decision to form a working group leaves everything up in the air. But one senior government official remarks that it will at least allow Japan to wait and see if Bush, a key backer of the SSC, is re-elected this year. **David Swinbanks**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

President Bush talks with Japanese Prime Minister Miyazawa.

TECHNOLOGY POLICY

South Korea commits funds

Tokyo

THE South Korean government has launched a major effort to become a technological power rivalling Japan and leading Western nations.

The government has set aside more than \$160 million (121,000 million won) in fiscal year 1992 for the first year of the 'Highly Advanced National Project', more commonly known as the 'G-7 Project' after the G-7 group of nations (United States, Japan, Canada, the United Kingdom, France, Italy and Germany) that South Korea hopes to emulate. South Korea's technology-related ministries hope to spend more than \$3,000 million on the project during the next ten years with a matching investment from industry (see *Nature* 354, 176; 21 November 1991).

The budget for 1992 and projections to 2001 (see table) show that South Korea plans investment in the development of next-generation integrated circuits, an area in which it already holds considerable expertise. Development of computerized manufacturing systems, an integrated services and data network, biotechnology, and next-generation transportation systems are each expected to receive support of several hundred million dollars during the next decade from both government and industry.

South Korea, after a period of rapid industrial growth after the Korean war, is

now faced with high inflation, large trade deficits and rising labour costs. The government says that the only way forward is to develop value-added high-technology products, but the population of trained scientists and engineers in South Korea is still limited, although growing rapidly. So the Ministry of Science and Technology, which controls the largest share of the G-7 budget, hopes to establish international research and development projects on which it is prepared to spend 5-20 per cent of its funds. The government also plans to allow foreign researchers to participate in national projects, in much the same way as Japan has recently opened up its government-industry research to foreign participation.

The G-7 project is the first attempt by South Korea to coordinate research by all the various technology-related ministries and agencies. In the past, lack of coordination has led to overlap of research investments, while individual budgets for each ministry have been too small for them to undertake large projects. Much of the research will be carried out by the Korea Institute of Science and Technology in Seoul and other institutes affiliated with the Ministry

of Science and Technology. But the new Korea Academy of Industrial Technology under the Ministry of Trade and Industry will also play a key role in coordinating the near-market research by government laboratories, universities and industry.

David Swinbanks

BUDGET FOR SOUTH KOREAN G-7 PROJECT

thousand million won

	1992	by 2001
Products Technology Development:		
Highly integrated semiconductors	15	365
Integrated Services and Data Network	20	250
High Definition Television	26	113
Electrical vehicle	3	72
Intelligent computer	3	115
Medicines and agricultural agents	8	136
Advanced production systems	7	257
SUB-TOTAL	83	1308
Fundamental Technology Development:		
Advanced materials	4	117
Next-generation transport systems	7	209
Biotechnology	5	241
Environmental technology	7	195
New energy resources	10	128
New atomic reactor	2	23
Human interface technology	2	40
SUB-TOTAL	38	954
GRAND TOTAL	121	2262

\$1 = approximately 750 won

Note above figures are for all government ministries and agencies but most of the allocations are for the Ministry of Science and Technology, the Ministry of Trade and Industry, the Ministry of Energy and Resources, and the Ministry of Communications

NIH wasted millions on computers

Washington

IN 1988, computer officers at the U.S. National Institutes of Health (NIH) contracted with IBM for a major computer system with a potential value over ten years of \$806 million that was meant to meet all of the institutes' computing needs. Unfortunately, those responsible for deciding to go with an IBM mainframe system failed to ask NIH scientists what kind of computing systems they wanted. Had they done so, according to a report from the US General Accounting Office, they would have discovered that many of the scientists favoured plain PCs and have no need for the "total system" for which NIH issued a contract.

The GAO report, which current NIH officials have not contradicted, records a priceless example of bureaucracy run

amok. Although the NIH have various computer advisory committees, the bureaucrats who chose the IBM mainframe system failed to consult any of them. Nor, according to the GAO, did they interview scientists because "they did not believe it was worthwhile to spend time surveying scientists," in part apparently because of a survey 20 years previously that "produced unmanageable results".

The total system NIH now has not only has excess capacity, but also includes one full-sized computer dedicated to backup and related needs. However, GAO notes that NIH mainframes (leased in earlier years) are connected in such a manner that one backs up another during a system failure. "Consequently, NIH mainframes virtually never fail and have been so reliable that NIH no longer keeps mainframe

failure data."

As to the scientists, who were heard from through an advisory body 4 months after the 1988 contract was signed, the availability of new software for personal computers rendered the need for mainframes for most scientific research outmoded. Besides, they said, the user charges for the multi-million dollar IBM system were too high.

So, while most researchers are happy with their PCs, NIH has wasted millions on a system is apparently does not need. GAO recommends that in the future NIH should solicit data on scientists' needs before buying them equipment they do not need and cannot afford. The current NIH leadership admits that is pretty good advice.

Barbara J. Culliton

NEWS IN BRIEF

Green scheme to open Asian markets

Washington

HOPING to open markets for US environmental companies in Asia, President George Bush announced plans last week to use the resources of 18 US agencies to match US experts and technology with Asian environmental problems. The agencies will sponsor trade fairs in Asia, travel to US environmental trade shows for Asian buyers, and environmental training institute in a Asian country to be named later. The US Agency for International Development will also start a biodiversity programme that will sponsor collaborative research on sustainable agriculture and conservation of genetic resources in Asia. Speaking in Singapore on 4 January, Bush promised that the initiative, known as the United States-Asia Environmental Partnership, "will be good for Asia's environment and good for American jobs." Environmental groups are concerned that the balance favors the latter. "It's mostly just the [Administration] agenda of increasing business opportunities," says Lawrence Williams, director of international programmes at the Sierra Club.

Chris Anderson

HUGO opposes Venter

London

THE Human Genome Organisation (HUGO) is contacting government officials and patent offices around the world, after deciding to oppose the attempt by Craig Venter, a researcher at the US National Institutes of Health, to patent more than

300 cDNA sequences. HUGO president Sir Walter Bodmer says that patenting cDNA sequences — fragments of genes whose function is not yet known — could lead to research groups working "in isolation and inadvertent competition, seriously slowing down research". HUGO is the second body representing genome researchers to oppose Venter's move, following the lead of the American Society of Human Genetics.

Nevertheless, NIH officials who bear responsibility for filing a patent application (Venter himself does not decide) are determined to press ahead even though they will set off what has been called "an enormous cDNA arms race (see *Nature*, 353:485). Critics of the patent application have argued that the filing is an abuse of the patent system, but NIH attorneys find the allegation irrelevant. Their concern is to convince the patent office that the cDNA sequences Venter has found have some utility (which is required before a patent can be granted) even though neither Venter nor anyone else yet knows exactly what the sequences mean.

Peter Aldhous

Institute for best and brightest

New Delhi

MAHARASHTRA state will soon set up an institute of advanced studies for exceptionally talented students. The new institute will be located in Bombay and offer an integrated five year course leading to a master's degree in basic sciences, with opportunities for students to work with leading scientists in national laboratories for their doctoral programmes.

The institute will admit 400 students each year. In contrast to existing government funded educational

institutions that reserve places for students from economically weaker sections of society, the proposed institute will enrol students exclusively on merit on the basis of state-wide competitive examinations.

Proposals to set up institutes of excellence for really talented students were made three years ago by the scientific advisory committee to the prime minister and the Indian National Science Academy. Now that Maharashtra has taken the lead, other states are expected to follow suit.

K.S. Jayaraman

Training for Egyptian scientists

Cairo

A COLLABORATIVE project between the University of California at Davis and the Veterinary Serum and Vaccine Research Institute here is the latest effort by the U.S. Agency for International Development (AID) to staunch the brain drain of African scientists. The agency has awarded \$1.9 million for a research project on a vaccine for rinderpest (from the German for cattle disease) that is meant to help in the transfer of recombinant DNA technology from the United States to Egypt.

During the next four years, 11 Egyptian scientists will study at Davis in the laboratory of Ethiopian virologist Tilahun Yilma who is director of the collaborative project. There they will learn how to make recombinant vaccine for rinderpest, which kills hundreds of thousands of cattle in Africa every year. Meanwhile, AID funds will be put to use building a modern molecular biology laboratory here which, it is hoped, will serve as a lure to attract the Egyptian scientists back home after training in California.

Jane Ferrell-Stevens

Support for Russia *et al.*

SIR — We write to support your suggestion (*Nature* 354, 339 & 499; 1991) that there should be an office in Moscow, and perhaps in other republic capitals in the former Soviet Union, to proffer information to researchers about assistance that may be available from elsewhere. This may be just what is needed in present circumstances.

You refer to the depth of the crisis in Soviet science. Economic recovery will no doubt be necessary to resolve the problem, but the capacity to innovate, an essential ingredient for recovery, is now itself at risk.

The immediate problem is especially acute. The former Soviet Academy of Sciences has disappeared. The Russian academy that replaces it promises well, but it is not yet clear what it can do. The hard currency budget of the Soviet academy was empty, so that the new Russian academy will not be able to provide the members of the laboratories it inherits with imported consumables, spare parts for repairs and other essential things, let alone travel abroad.

Funds are also urgently needed to transform present military research into work for civil industry. There is a real danger for peace that yet more defence workers, in the absence of alternatives, will be attracted abroad, using their skills for others. Self-interest as well as altruism makes it necessary for the international community to help former Soviet scientists to stay at home.

That is why we hope that the help you recommend can be organized promptly. May we also suggest a commercial means of keeping research capacity alive? There is a great deal of Soviet science, notably in mathematics, pure and applied. Some of it, in electrochemistry for example, leads the world. Many scientists are seeking means of working effectively and even feeding their families by taking on contract work specified from abroad. But they need help in undertaking market-led research.

The Russian Commodities and Raw Materials Exchange appreciates that its long-term prospects depend on continued and appropriate innovation that will benefit the economies of the new Commonwealth of Independent States. The exchange therefore plans to use some of its income, based on a levy on trade, to support applied research by talented people. This will be spent on salaries and facilities, and will not be used to prop up the old centralized system. We would be very glad to hear from any Western partner willing to propose concrete applied research projects.

As we see it, the partners would set up a joint stock company to do the work and

to assist both Soviet researchers or research organizations and Western interests wanting to commission research. As the proposed law on intellectual property becomes established, it is intended that the profits from successful innovations should be divided between researchers, customers and financial backers on a fair and agreed basis. Thus the initiative should become self-financing and, in the medium term, profitable.

This will not in itself help basic research, which cannot attract funds in such a way. Yet without basic research, neither applied research nor good teaching can flourish. The notion that some ten per cent of research budgets should be spent on basic research is sometimes mentioned. Because pure research is especially vulnerable to market forces, we plan to set aside a greater proportion of what we provide commercially for distribution to the most promising basic research. We seek Western partners willing to do likewise.

JOHN GODFREY

41 Lawford Road,
London NW5 2LG, UK

ALEXANDER M. HAZHAKYAN
(Manager & Executive Director)
*Russian Commodities
& Raw Materials Exchange,
Frunenskaya nab. 30 Pav. 3,
Moscow 11914, USSR
(Fax: 095 262 5757)*

Ancient error

SIR — One of the most absurd and persistent misuses of words in science is the use of the word 'ancient' to describe species that flourished millions of years ago. Thus, trilobites, dinosaurs and the creatures of the Cambrian Burgess Shale are all described as ancient. The earliest of the three classical geological eras is known as the Palaeozoic, the era of ancient life.

The truth is that organisms that lived long ago are 'young' relative to those alive today. 'Life', like the world inhabited by living things, is older now than it was millions of years ago. Thirty years ago I was young, not ancient. We do not refer to our youth as the ancient time of our lives, but the early time. Palaeozoic times were, surely, the age of early, not of ancient, life. Similarly, we talk of 'Ancient' Greece and 'Ancient' Egypt, as if those countries used to be old and have since grown young. The opposite, of course, is true. Three thousand years ago, Greece and Egypt were young countries and, like everything else in this world, have aged considerably since their early days. The Greece and the Egypt of today are ancient. Yet we

persist in thinking of people who lived long ago as the 'Ancient' Greeks or the 'Ancient' Egyptians, when the adjective 'Young' or 'Early' would be more suitable. One constantly hears remarks like "the Babylonian civilization is of great antiquity", but what we really mean is that we have known about the existence of the Babylonians for a long time. This makes them 'ancient' in our eyes. The same applies to dinosaurs and trilobites. The rocks in which we find the fossils of these animals are, indeed, ancient, because they have been there for a very long time, but the animals themselves are not ancient in any sense at all. They died out a long time ago, when the Earth was young.

Why do we persist in misusing the term in this way? I have a suspicion that terms like 'old' and 'ancient' are used as expressions of awe and reverence, since we normally view the subject matter of palaeontology and archaeology in that light. Perhaps we cannot look upon the great learning of the 'Ancient' Greeks or the majestic 'terribleness' of the dinosaurs as young things. Perhaps we feel that such magnificent and awesome things cannot possibly be young — Aristotle and *Tyrannosaurus rex* are venerable beings.

But, it may be argued, what does it matter whether we call early life 'ancient' or not? It can be said that it is merely a convention and that it is sheer pedantry to split hairs over this matter. I do not believe so. Words can play subtle tricks in the mind and any scientific word that might lead us astray in our thinking must surely be replaced with one that is more suitable. I am inclined to think that 'Early' is the best alternative.

ZAKARIA ERZINCIOGLU

Department of Zoology,
University of Cambridge,
Downing Street, Cambridge CB2 3EJ, UK

Motion sickness

SIR — Daedulus is correct in noting (*Nature* 354, 438; 1991) that "motion sickness makes no biological sense".

What makes even less sense is the near-imperviousness to motion sickness of children. My two-year-old daughter, for example, loves to be spun on a barstool to the point of nystagmus. Rolling a hundred feet or more down a grassy hillside is another of her ideas of having fun. A stroll through any amusement park will show that this is generally true.

What can account for this? Perhaps the solution to motion sickness is to induce stronger neoteny in the development of our semicircular canals.

TIM GORSKI

2702 Laurel Valley Lane,
Arlington, Texas 76006, USA

Leaking gas and greenhouse

SIR — Max K. Wallis (*Nature* 354, 428; 1991) claims that accelerated replacement of gas mains would be a cost-effective means of reducing greenhouse gas emissions. Unfortunately, his argument is based on a series of erroneous assumptions about the amount of gas leakage and the costs of replacing mains and services.

The UK Department of the Environment estimates show coal-mining as the major source of methane emissions. Its latest estimates, for 1989, show that gas leakage from the distribution system accounted for 340,000 tonnes of methane (*Hansard*, 11 November 1991, Written Answers, col. 390). Using Wallis's conversion factor, this comes to 170 million therms — very different from the 908 million therms "avoidable leakage" he assumes.

The length of pre-1969 mains still in use is 100,000 km, not 140,000 km. Wallis was told this by British Gas in August 1990 and November 1991, and I fail to understand why he uses a much higher figure which he knows is not accurate. It is also a matter of fact that a considerable proportion of the 100,000 km was not laid with hemp-lead joints, because this technique was discontinued long before 1969. British Gas also denies the implication that gas conditioning to reduce leakage is "hardly effective" for low-pressure mains. Considerable amounts of money are spent on this activity which forms a major part of the leakage control programme.

Furthermore, Wallis has misunderstood the figures he has taken from the company's annual report and accounts for 1990, and in doing so has grossly underestimated the true costs of replacing both mains and services.

His assumptions are thus very far removed from the true figures. They do not add to the well-established case, accepted by responsible gas suppliers such as British Gas, for retaining a programme of replacing mains and services. This is justified by the need to maintain public safety and security of supply. Over the past ten years, the number of serious gas explosions has

halved, while the amount of gas consumed has increased by more than 15 per cent; and British Gas has relaid 30,000 km of mains and replaced more than 4 million services, mostly using welded polyethylene pipe which is virtually leak-free.

There is no need for arguments about the greenhouse effect based on erroneous assumptions. The maintenance of a safe, reliable and economic system of gas supply will always require that gas leaks be kept to a minimum, and that the replacement of mains and services should be a continuous planned activity; but there is no justification for a crash programme of total replacement. British Gas intends to continue with its massive programme of mains replacement, to keep the public informed about its activities, and to counter the greenhouse effect by two means that are known to be effective: encouraging customers to conserve energy, and supplying gas in place of fuels that emit much higher levels of carbon dioxide.

R. HERBERT

*British Gas,
152 Grosvenor Road,
London SW1V 3JL, UK*

Dirty habits

SIR — James Vickers (*Nature* 353, 103; 1991) mentions the risks inherent in bottle-feeding where clean water is scarce and infection rife. M. F. Goldfarb stresses the problems resulting from mixing milk powder with poor-quality water. These statements reflect the prevalent dogma that infected water or baby foods are a major cause of infantile diarrhoea, that is, the problem is infected oral input. I have just received circulars from two charities: WaterAid states that dirty water kills 10 million babies and small children every year; CARE states that in a year 3.5 million children under five could have perished, victims of just one cause: diarrhoeal disease spread by dirty water.

Over many years, I have been checking published data after suspecting that airborne infection via latent otomatoiditis with secondary pulmonary and intestinal infections might be the main cause of postneonatal infant deaths. In *The Lancet* in 1981¹ I claimed that there was no relation between contaminated milk or water supplies and infant gastroenteritis. This was not challenged². In the *Journal of the American Medical Association* in 1986, I claimed that the crucial factor was how infants were fed, not what they were fed. One person disputed this³, but apart from making the obvious point that bacteria are ubi-

quitous in developing countries, no convincing experimental or observational study directly linking contaminated food with infant diarrhoea was cited², or has been produced since.

In view of the major health, economic and political implications of this dogma, it is unhelpful to keep asserting it, as it then becomes a self-evident truth which no one thinks even to test. It is a new form of cultural imperialism — or maybe the emperor has just got new clothes?

A. G. GORDON

*32 Love Walk,
London SE5 8AD, UK*

1. Gordon, A.G. *Lancet* II, 151 (1981).
2. Gordon, A.G. *J. Am. med. Ass.* 257, 3362 (1987).
3. Holland, B. *J. Am. med. Ass.* 257, 3361 (1987).

Quality meetings

SIR — Last July, the International Programme Committee of the International Union of Physiological Sciences (IUPS) met in Glasgow, to prepare the union's 32nd congress. The meeting was well organized and productive, anticipating a success for the IUPS meeting in 1993.

But a question arose which in my experience, is not limited to a given scientific area or to a working group: how is it possible to assure the highest scientific quality of a 'world congress' and at the same time have a fair geographical representation? This is particularly important for scientists working outside North America and Western Europe who want to share their experiences with their colleagues. Taking the IUPS as an example, about 4 per cent of those in the World Directory of Physiologists are Latin Americans, a figure that roughly agrees with the written production of Latin American scientists in the life sciences (*Current Contents, Life Sciences*, Issues 19 & 20; 1984). However, fewer than one in a thousand lectures and symposia speakers came from this geographical area during the last meeting in Helsinki (Proc XXXI IUPS Congress, 1989).

A simple and feasible answer has to embrace both aspects of the question. First, quality: invited speakers should be selected from among the best, taking into consideration as an objective index their publications.

Second, a fair geographical distribution among the potential speakers who publish in leading international journals: why should scientists working outside North America and Western Europe not have the opportunity to discuss the same results as are accepted for publication?

CARLOS LIBERTUN

*Laboratory of Neuroendocrinology,
Instituto de Biología
y Medicina Experimental,
Obligado 2490,
1428 Buenos Aires, Argentina*

Hit or miss?

SIR — Fortunately for us all, the encounter described by Scotti *et al.* (*Nature* 354, 287; 1991) was not, as characterized by the authors, a "near miss of the Earth by a small asteroid" but was, rather, a near hit.

STEPHEN E. SCHWARTZ

*Brookhaven National Laboratory,
Upton, New York 11973, USA*

Uncertainties of climate change

SIR — To resolve the uncertainties enveloping global climate change, the United States and, to a lesser extent, Europe have embarked on an extensive research effort. We believe that these research programmes will do little to provide a solid scientific foundation for policy decisions in the next decade.

Marginal improvements in current models, or additional runs of the global circulation models, will not resolve the fundamental uncertainties paralysing international negotiations. Research is at present focused on predicting the extent of climate change, a question that is unlikely to be resolved for at least another decade. More likely to be resolved and of greater importance in determining policy are the cost of abatement and the effects of climate change. In our view, scientific research is destined to remain largely irrelevant to political decisions about global climate change as it is not designed to produce answers when needed by policy makers.

These problems were the downfall of the decade long, half-billion-dollar National Acid Precipitation Assessment Project (NAPAP). Its important scientific findings had virtually no effect on policy and legislation. Research on the greenhouse effect is following the same path. If science is to be relevant to social decisions about global climate change, the programme's management and focus must be changed. The focus should be on informing policy decisions over the next decade or two, not on abstract research. Fundamental research is needed, but the agenda must be structured to answer crucial policy questions, not simply to advance knowledge.

The United States is spending \$1,200 million on climate change research this year, but only 25 per cent of these resources are focused on the core issues. Even these programmes were generally

designed to answer other questions. If scientific research is to inform political decisions, administrators need to exercise tough control. An integrated assessment is needed to identify the priority research and coordinate individual projects.

In our judgement, several independent assessments should be undertaken in parallel. Integrated assessments can spot the critical gaps in the current research agenda, discover research that isn't on target or is wasteful, and detect the mismatches between the inputs that each group is expecting and the outputs that will be produced by other researchers. Global climate change is too important to repeat the same mistakes that crippled the \$500-million NAPAP programme.

LESTER B. LAVE

HADI DOWLATABADI

GREGORY J. MCRAE

M. GRANGER MORGAN

EDWARD S. RUBIN

Carnegie Mellon University,
Pittsburgh, Pennsylvania 15213, USA

The grant racket

SIR — Science advances by the independent thinking of nonconformists. Today, nonconformists have a hard time in US academic research. For instance, PHS (US Public Health Service) funding is based on peer review, but only about 15 per cent of all approved new requests are now funded¹. Each request takes months to prepare and submit; revisions can add years of delay. Pressure to get funding supervenes the drive to test new ideas. Success is determined by mastery of idiom more than by scientific vision. Departure from accepted views guarantees derailment in peer review. Tight money imposes conformity where independent thought is required.

Behind this are scientific and grant administrators who contribute little to science and often impede it. While grant money now pays 'soft' salaries and overhead, universities pay administrators 'hard' money and let them control endowments. 'Blue-collar' scientists who founder in peer review risk more than 'white-collar' administrators who do not even work in the laboratory. The system favours the wrong group because administrators, not bench scientists, influence policy: foxes guard the chicken coop. To cure this, Congress should set the following PHS policies.

(1) *Harden 'soft' money.* Limit PHS research support to \$70,000 per year per investigator above salary until all approved applications are funded. Congress recently gave pay rises to upper-

grade civil service scientists (and to itself) when productive laboratories are dying and only 15 per cent of good new ideas are being studied. Instead, let PHS augment salary to, say, \$50,000 per year maximum and exempt those making more than that from PHS support. As an incentive to get outside money, let any extra salary be used for research.

(2) *Set PHS guidelines on technology ownership and transfer.* Most universities claim all inventions conceived within their walls. However, the inventor's share of licensing and royalties ranges from zero to half. The even split (i) stimulates inventiveness, yet bypasses the stifling academic bureaucracy; (ii) deters inventors from walking off with a super idea; and (iii) frees salary money to less lucrative research areas. This does not sully pure research: fiscal strain is killing basic research, and this may preserve it. New approaches to technology transfer are also needed.

(3) *Soften 'hard' money.* Today, the award rate for approved individual investigator grants is in sharp decline. Yet 33 per cent of NIH funding is 'indirect costs', the fastest growing category¹. The increase is largest in private universities whose endowments once paid overhead instead of paying for money-acquisition machines and administrator slush funds. PHS should let endowments pay 'indirect costs'. The increased funds could nearly double the diversity of independently run PHS projects.

(4) *Researchers should sign papers only when they do the work.* Before US science became big business, authorship was obvious. The feudal structure of academic institutions encourages ambitious bureaucrats with fiefdoms to exact credit as fealty for resources. This stimulates fraud. Perhaps we need a 'secondary authorship' line for technical services. But when administrators can no longer exact credit for science they did not do, it will be easy to fix the blame for bad work. This will inhibit fraud without data auditors in trenchcoats.

PHS will not take these initiatives voluntarily, because it is a bureaucracy. But, if Congress forces it to, private funding agencies will agree: they often limit overhead to little or nothing. Fiscal reform will show if science is really underfunded in the United States. Support many, small laboratories, not a few big ones; protect rare breeds in academic research, as in wildlife conservation. Trade each \$100,000 per year fox for two independent scientists to spur science and squelch fraud.

PETER M. ROSS

Box 7022,
New York, New York 10128-0010, USA

1. Extramural Trends, FY 1980–1989. Information Systems Branch, Division of Research Grants, National Institutes of Health.

Good heavens!

SIR — I was alarmed to read in your editorial on Olbers' paradox (*Nature* 352, 554; 1991) that "the brightness of receding stars will be diminished by the recession". While I have observed numerous effects on science of the current economic downturn — from cuts in research council budgets to increased applications for PhD places from graduates squeezed out of the job market — I had hitherto assumed its consequences to be confined to the merely terrestrial.

PETER A. LUND

School of Biological Sciences,
University of Birmingham,
Edgbaston, Birmingham B15 2TT, UK

Handicaps to British 'innovation'

Alexander Kennaway

Britain urgently needs to return to traditional, but now abandoned, values in science. Scientific administrators and EC directorates should revise their policies in the light of Japan's experience.

THAT Britain is good at producing ideas but bad at successful commercialization is a refrain heard down the decades. Britain spends per capita, from public funds, more than many of its more successful industrial competitors, and its leading academics demand more as a means of rescuing not merely the country's industrial base but also its performance in scientific research which, they assert, is beginning to slip well behind the leaders. The shortage of money and the general atmosphere of retrenchment in universities and research institutes is so depressing that it is feared that many promising young people will prefer to work abroad. Others, especially politicians and businessmen, ask for 'value for money'.

Increasingly apparent is the growing intention to control scientific and technical innovation by submitting applicants for funding to disciplines which are appropriate to the management of projects and to the later stages of development and application of novel basic ideas. Every official grant-giving body in Britain and the European Communities pays lip-service to the notion of doing innovative, curiosity-led research, but requires the applicant to state in advance what capital and revenue expenditure is envisaged for some years. The UK Department of Trade and Industry officially requires those in its research laboratories to put time limits on research projects. Indeed, some scientists have felt obliged to put in their 'mission statements' (a current fashion) that they will complete 95% of their research projects within the stated time.

These procedures are inappropriate to research and counterproductive to genuine innovation. Perhaps they stem from an innate feeling among those without the background, ability or flair to be creative in science that scientific researchers are not to be trusted to select their own lines of work in a responsible manner. It is as if there remains in the psyche of the adult English middle class who form the country's establishment and business leadership a memory of the 'mad scientist' who inhabited his childhood reading in comics and thrillers, and who must be trammelled by constraints from the world of business and its disciplines. The performance of British industry, seduced and dominated as it is by the norms of the city of London,

particularly (but not only) since the Second World War, has hardly been inspiring; its products have become less competitive and its share of world markets, even at home, is shrinking fast. For grant-awarding bodies to impose their methods, especially of financial control with its impatient demand for quick returns, upon innovation, whether scientific or technical, is not only unjustified but is killing the goose with its potentially golden eggs.

It is also counterproductive for politicians to impose business and pseudo-business methods on the academic and research worlds. The time has come to challenge and to abandon the vaunted

path to acceptance. The results are that grants are made to those who possess those skills but who may not be good researchers. Furthermore, the good researchers are diverted from their most productive work. Some indeed become so frustrated that they give up asking for money, to the detriment of their work.

Traditionally, Britain has relied heavily upon its universities for fundamental research. The clarity of blue-sky research is now being clouded by the urgings of government that universities derive, on a growing scale, income from research contracts and indeed from direct technical service to industry. This policy has two weaknesses. The first is the distortion towards day-to-day work that diverts from fundamental work unlikely to be done elsewhere. And second, companies are in danger of thinking of universities as a cheap extension of their own laboratories. Both government and industry are accustomed to paying much smaller overheads than the full absorption cost of university work. British industry argues sometimes that it already pays through taxes and that it should not pay twice, but overlooks the fact that government funds do not cover the full overhead of teaching facilities, let alone that allocated to research. There can be no objection to doing some such work; but the proper policy is a question of balancing both volume and quality. This process promotes useful links at the working level between academics and industry, promotes technology transfer and provides practical examples that illuminate teaching. It also brings badly needed money into the universities.

The successes of British innovation have been obtained under totally different cultural regimes to which it is almost certainly not possible to return. However I am confident that the present arrangements and climate for innovation can be improved upon. Let us begin by studying the situation pertaining in Japanese science and techno-commercial innovation. Britain has much to learn from these studies but there are also severe limitations on what can be done because of the poor condition of purely British manufacturing industry.

Japan's experience

As everyone knows, the Japanese were prostrate and literally starving at the end

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

M. Peckham

British basics — tools for innovation?

'customer-principle' advocated by Lord Rothschild and practised by government institutions ever since. The principle has some validity in projects which have a short-term commercial target, especially where a development department needs some more basic work than it can do itself, but it can have only an inhibitory effect on true research.

It is becoming increasingly difficult for truly creative people to explore the frontiers of knowledge as they are perceived today. The whole panoply of regimes for the management of national research leads not to creativity but to safe, pedestrian projects which will be accepted by the committees and peer-group reviews in surrogate government grant-giving bodies. Indeed, one might observe that a significant amount of time and ingenuity is being spent by people on writing proposals in a form that smooths the

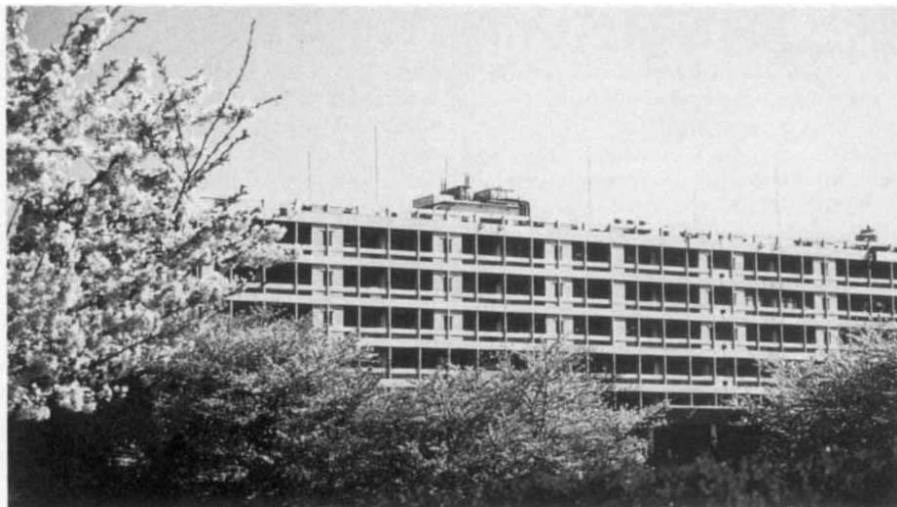
of the Second World War. The Japanese people knew that they had to resume the efforts to modernize and industrialize their country that had begun after the Meiji restoration in 1868, but survival took first place. With hardly any natural resources, exports were essential to pay for import of basic necessities. The Japanese began with basic engineering, much of it based on imitation of good foreign products. Although Japanese products initially did not command much respect abroad, after 15 years or so they began to be recognized as world leaders. This success, which has developed to the point that Japan is a world economic and technological leader, has been due to many factors which can be summed up quite shortly.

The Japanese studied, absorbed, adapted and applied the best practices from elsewhere to their engineering industries. Their business objectives, as stated by company president after president, however, are their own — to identify sectors of the world market which they can serve uniquely well, to do so with ever better standards of excellence, and in so doing to earn a good income to enable the company to survive and grow in the long-term. Included in the company are its employees, suppliers and other business associates.

The pursuit of excellence is enabled by a management style totally different from that of Western business schools, which the Japanese have studied and often rejected. Japanese managers identify their job as providing leadership, motivating and training people to perform to ever better standards, where necessary showing them how. In other words, the young and middle managers are leaders in a field of which they are already the masters. They pursue personal career ambitions through service to their team. They do not seek to maximize their financial performance in the short term by cutting long-term expenditure on training, capital investment, maintenance and, above all, on innovation.

Innovation policy

Innovation policies provide the crucial key to success in product development and innovation, the fuel for commercial and financial success. Eighty per cent of all funding for science and technology research and development in Japan is funded by private companies, government agencies supplying the rest. Slowly, these firms built up their expenditure; a wide range of engineering firms is now allocating about eight to ten per cent of their total income from sales to research and development. Of that huge sum, about ten per cent is allocated to purely speculative, non-targeted research. Such



RIKEN laboratories — a model for others to follow?

sums are rarely paralleled in Britain or the United States except historically by the major chemical and pharmaceutical companies: even there, regrettably, the pressures exercised by the financial world are reducing expenditure in the move towards short-term results.

It is not just the sums of money that provide the Japanese with the fuel for commercial innovation and for basic research, but the outlook and policies that lie behind the money and the way it is spent. Japanese companies to a far greater extent than elsewhere are determined to be intellectually independent, consequently research workers in firms making cars, consumer and industrial goods as well as consumer electronics are studying new materials at the atomic level as well as cognitive science. Their research directors know that such work, if successful, may not produce a commercial and financial return for a long time — they are prepared for such ideas to take 10 years before they emerge from the laboratory and for up to 30 years before the project yields a net financial return. Japanese research leaders, whether in industry or in independent institutes such as RIKEN (see below) possess to the full the ability to encourage new science and at the same time to realize that if it were to become successful then it would result in novel commercializable, useful and profitable technology and products.

Japanese companies encourage young employees to explore new frontiers of knowledge; indeed in many firms the proposals for new research come from the bottom level of graduates and not from the top down. Section leaders and their superiors have contingency funds to allow exploratory, pilot work to be done without going through formal systems of submitting and approving the work by research committees. If the exploratory work seems promising then it is included in next year's budget. Budget decisions are often taken in informal evening

meetings of junior and middle managers of research laboratories. The director attends but, so some of them told me, they rarely differ from the consensus of their juniors.

To an Anglo-American research director, the idea of funding a speculative research programme which may not pay its way for 20 or 30 years would be a dream; in Western society background research projects in firms are liable to be stopped if they do not pay back in less than 5 years. The Japanese can afford this luxury for two reasons: first, they have been at the leading-edge of research for about 20 years and the fruits are already visible in company profits; and second, the operating divisions are fully responsible for product and process innovation and they have been successful at applying new science and technology to produce improved and new products and manufacturing systems with a maximum lead-time of 2 to 3 years. Japanese companies recognize that as products become more complex it will be more difficult to sell products that are merely gimmicks and that have simple facelifts. This implies that lead times for new products will inevitably become longer. One method to reduce this lead time, called simultaneous engineering by Nissan, assembles a team including a specialist from every branch of the company from sales to product design, styling, manufacturing and after-sales service. In this way, employees rapidly become generalists. At Nissan, for example, 18 months has been taken off the 5-year cycle time for the introduction, from conception, of a completely new car. In the West, such methods have been the common practice in the multinational chemical industries for decades.

The Japanese are increasingly recognizing that it is too expensive to innovate basic products entirely alone, so they search for competent partners, often abroad. One research director told me that his company found that the best way

of determining whether a potential foreign collaborator would 'fit' was to exchange visits for research workers to each other's basic research laboratories — a refreshingly open policy.

The quality and innovation that has been invested in the high-profile product areas such as cars and consumer electronics could not have taken place had Japanese companies not been able to rely on the outstanding quality and innovation that has taken place in fundamental engineering. General engineering also provides by far the biggest export earner from Japan, outpacing both cars and consumer electronics by 5 and 10 times, respectively.

Basic research

The ability of the Japanese to design and produce some of the world's best instrumentation also plays a significant role in the leading position occupied by much of Japanese science. There is a significant interplay between engineers and scientists so that the engineering products respond to the most discriminating market in the world, the Japanese customer, who in turn can use outstanding tools to do excellent science in the microspheres of materials, genetics and X-ray astronomy.

Japanese government institutions involved in science pay great attention to the opinions of industrially based scientists and technologists when they come to consider which scientific targets should be chosen and treated as national projects. Industrialists take comfort from the national consensus and are reinforced in their choice of programmes, which are carried out in their own laboratories.

Every department of the Japanese government is keen to obtain significant budgets to support scientific research in its own field. The Ministry of Education controls half the public funding for science which is largely carried out in postgraduate institutes of Japanese universities. (A familiar Western model might be Princeton.) The difference is that in Japan there are several hundred universities and the Ministry of Education seeks to provide some funds for almost all of them, thus the scientific community complains that these funds are spread too widely and that the sum granted to each applicant is rather small and therefore inefficiently used. This is notwithstanding the very large sums invested in, for example, the fusion research laboratory attached to Osaka University.

Consequently, the Japanese scientific community searched for other ways to fund science. It found two methods which have similarities. The science and technology agency, which reports

directly to the cabinet, directly funds the RIKEN laboratories, of which there are about 50 in one building. (The acronym stands for the Institute of Physical and Chemical Research.) The laboratory does not fund or look for good projects but seeks out outstanding research leaders and who are themselves highly innovative. Such people tend to be selected from young PhDs who may work in government institutions, universities or private companies; they are asked for whom they would like to work, those who receive the most support are invited to work at RIKEN and are funded for 5 years. They are told that they have total freedom to choose their projects and that they have to find between 5 and 10 young research workers. The group leader is given basic funding but once he has settled on a line of work, he is given much more and he may invite an equivalent number of workers on a short-term basis from Japan or from abroad. The laboratory is deliberately multidisciplinary, the basic facilities in many branches of science are provided but no boundaries exist to prevent groups in different areas and disciplines from working together. After 5 years the team is disbanded. Although it is an object of RIKEN to see that at least a third of its 50 groups are productive (that is to say that useful results emerge from good work), its primary objective is to turn out highly motivated, experienced researchers who will go on to work elsewhere.

The second set of programmes, called ERATO, operates similarly in most respects except that the work is done not in a special institute but in universities. The programme is identified by the name of its leader rather than by project title. Both RIKEN and ERATO programmes attract workers seconded from firms as well as from academic life. One of the most striking features of Japanese science is that it strives to look beyond the narrow confines of a specialist discipline; indeed it strives to broaden its approach to embrace humanities and general culture. Everyone in authority says that they have no difficulty in persuading the taxpayer to fund pure science, even in the expensive world of astronomy, adding that this fulfils a basic curiosity to find out the origins of life and the Universe.

There is no attempt by the authorities to produce 'tidy' schemes of the kind that would appeal to British bureaucracy; the Japanese have not abolished the programmes and methods used by government departments, they have simply decided that alternative means lead to valuable results. They also say that indices of performance and productivity such as scores in the *Citation Index* are useless to measure good science, adding scornfully that they are loved only by politicians and

bureaucrats.

Conclusions

The Japanese experience demonstrates the value of returning to traditional but abandoned British values in science. Basic science and the first steps of engineering innovation cannot be 'managed'; it is better for the researchers to take responsibility for their own work, to see that ultimately a fair proportion of it will be useful. It cannot be right to impose so-called business methods, rules of thumb by accountants and illusory feasibility projections. Choice by personal qualities rather than by project management and proposals clearly is more likely to produce good work. Researchers have to be 'polyvalent' — multicultured rather than blinkered.

Short-termism is the death of innovation and much else in British and US industry. As a result, manufacturing industry is so weakened that it cannot afford even the investment in applied development work that has made the Japanese rich. Consequently it cannot be expected, as the British government hopes, to support basic research. Much of present expenditure on what passes for research and development in Britain is devoted to military uses; not only is this work not to be classed as true research but the new international realities will force a diversion away from this to more useful spheres. Therefore British research must increasingly look abroad to foreign and multinational organizations for collaboration and funding. This is already happening; at Imperial College, London, for example the contribution from purely British industry (excluding multinational companies) is about fifteen per cent of total research income — much of that in defence-related projects. It also receives about the same from UK charities, while support from abroad has grown to eleven per cent.

It is probably too late to create wealth by injecting novel science and innovation into British-owned firms; the country cannot pull itself up from the present as the Japanese did in 1945. The future for Britain lies increasingly in the establishment of foreign factories and research laboratories. This is a stated objective of the Japanese, who say that a local subsidiary is not truly British (or German) unless it possesses not merely a technical service and development laboratory but also fundamental research laboratory. Such institutions will provide employment for young researchers, who will also have to broaden themselves if they are to have a good career. □

Alexander Kennaway is in the Department of Mechanical Engineering, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK.

Is molecular biology yet a science?

The great successes of the past few years suggest that living processes consist of well-ordered events executed under strict control, but a few numbers would give a different and more fuzzy impression.

NOBODY can now shrug off the triumphant pervasiveness of molecular biology, which touches everything in the life sciences and which also affects the character of research. There is now an army of people called molecular biologists whose published papers are innocent of references to whole plants and animals and which may have little to say about their physiology either. For these people, experimental data may consist largely of what are called, in the trade, 'gels' — perhaps simply strips of filter paper allowing the comparison of the molecular weights of large molecules with those of similar materials by the relative rate at which they move under the influence of an electric field. Such techniques are deliciously simple, at least in skilled hands. It is marvellous that so much continues to be learned from them.

The other side of this coin is that molecular biology seems well on the way to becoming a largely qualitative science. The notion that science hangs on measurement seems to have been diminished (although, to be fair, electrophoresis is a measurement of a kind). For some purposes, that may not matter. If the objective is simply to tell the difference between the normal version of a gene and that which appears to be responsible for an inherited disease, then it may suffice to isolate fragments of each (by electrophoresis, for example) and then to determine their nucleotide sequences.

The worry is that, in the long run, strictly qualitative information such as that will not suffice. If some intermediate goal in biology is an understanding of the functioning of an entire cell, it is unthinkable that it will be attainable without quantitative information about the abundance of the component molecular species. But it is also possible that many subcellular processes on which attention is now concentrated will not be intelligible without new quantitative information.

To what extent, for example, is the timing of the transcription of a particular gene determined by the concentration of the particular protein molecule (the 'transcription factor') that must bind to all-purpose RNA polymerase before transcription can begin? And exactly how efficient are the processes by which membrane and other proteins migrate to the places in a cell where they 'belong'? If, in the development of an organism, the emergence of different parts is determined by the diffusion of molecular morphogens,

what concentrations of them suffice for normal development?

In the biochemical economy of a cell, such questions are far from irrelevant. If, for example, the transcription of a gene requires a protein specific to that gene, and then the formation of a complex between that protein and a suitably placed molecule of RNA polymerase, it is improbable that the pre-synthesis of a single molecular copy of the initiation factor will suffice. But the synthesis of more than a single molecule may be, strictly speaking, a waste of free energy. Is it not crucial to the understanding of the cell to know how great is the implied inefficiency?

In those bacterial systems in which cells are spontaneously transformed from one form into another, it seems that the accumulation of the transcription factor necessary to transcribe the genes active in one stage may also be the trigger for the transition to the next stage. The relationships between the regulatory processes concerned are intricate and, in their own way, elegant. But those concerned with them seem generally to be more excited by the intricacies than with the definition of what concentration of this or that element of the regulatory apparatus is needed for orderly development.

That is not only a misfortune in its own right, but part of the reason why those seeking explanations of, say, the onset of cell division persistently look for qualitative triggers when quantitative triggers — such as the accumulation of unwanted products of previous stages — may be a large part of the answer. More generally, a greater concern for the quantitative features of gene control might throw light on phenomena such as that in which children produce both fetal and adult haemoglobin for some months after birth, which could well be relevant to the understanding of maturation.

Curiously enough, even when molecular biology derives from fields in which there is a strong tradition of measurement, neurophysiology for example, attempts to make arguments quantitative appear to be neglected. To be sure, it is possible, using standard thermodynamic arguments, to relate the membrane potential of a neuron to the ratio of the concentration of potassium and other ions inside and outside the cell, but what does that imply for the rate at which such a cell must generate ATP so as to maintain the polarized state, let alone to

recovering from depolarization? The sums can obviously be done, but are their implications explored?

Even the textbooks are not enlightening. For instance, *Molecular Biology of the Cell* (Alberts *et al.*) and *Molecular Cell Biology* (Darnell, Lodish and Baltimore) have impeccable accounts of how mitochondria convert the energy of the respiratory cycle into ATP by maintaining a gradient of hydrogen ions across the inner membrane, but neither goes so far as to describe even back-of-the-envelope calculations to suggest what quantities and concentration may be involved.

There are two important consequences of the neglect of quantitative considerations in molecular biology, one of which is psychological: so long as people search for (and continue to find) qualitative explanations for phenomena in cell biology, they will give credence to the view that the average cell is just a bag filled with molecular switches that exist to be turned off and on as the appropriate molecular actuators make their appearance in some predetermined sequence. That helps to give the impression that the reductionist agenda would make the description of living things neatly cut and dried, which could not be further from the truth.

More seriously, the neglect of quantitative considerations may well be a recipe for overlooking problems inherently of great importance. Enquiries as different as the search for the origin of life and the understanding of the efficacy of the secretory mechanism in cells require not just the knowledge of what molecular processes are energetically possible but also some understanding of the concentrations in which participating chemicals may realistically be found. And there is every likelihood that, then, the dynamics of these processes would significantly limit the range of what is otherwise possible.

So should molecular biologists mend their ways, resurrecting the Law of Mass Action (now conspicuous by its absence from what they publish)? That would be to ask too much when there is so much to be learned from qualitative relationships. But it would be a worthwhile precaution against the quantitative days that lie ahead that people should make sure that published data are capable of quantitative interpretation by those who have the zeal for that. As things are, that is far from true.

John Maddox

The Galapagos that were

Hampton L. Carson

EXCITING new geological evidence, described by Christie and colleagues on page 246 of this issue¹, shows that a series of drowned islands (seamounts) occurs east of the existing Galapagos islands. The oldest seamount, 10 million years old and 2,500 m deep, lies 700 km east of an ancient fixed hotspot where magma wells up from the Earth's mantle. Lying on the plate above it at present is Fernandina, the youngest Galapagos island. Both the present islands and the old seamounts sit on the Nazca tectonic plate, which has been drifting eastward across the hotspot for about 25 million years².

Oceanic islands such as the Galapagos group began as sterile masses of cooled lava. Some such geological events are very recent and occur in areas remote from any continent, so that colonization by life is rare. If a unique species evolves on such an isolated island, the species must be no older than the age of that island. Using molecular genetic techniques to measure genetic distance, and employing the concept of a molecular clock, an estimated divergence time relative to a second species may be obtained. If this time is greater than the geologically determined age of the island, it is assumed that the divergence of the species occurred before the formation of the island.

The findings of Christie *et al.* thus provide fuel for debates about the age of the Galapagos biota. No present Galapagos island is older than about 3 million years, and some authors consider this time to be the maximum over which the exuberant evolution of the many unique Galapagos life forms has taken place³. Twelve diverse species of Darwin's finches fit well into this hypothesis, the protein-based genetic distances between them being so small that divergence could have occurred in as little as a million years⁴. On the other hand, a much larger genetic distance has been found between the marine and land iguanas. From immunological data it seems that the two lineages must have separated 15–20 million years ago⁵, and the existence of the drowned islands suggests that the divergence might have taken place on such islands before they sank beneath the surface. Another explanation is that the large genetic distance between the iguanas might be due to their origin from two different ancestral stocks on the South American mainland, although no such source populations have been found.

Certain quasi-neutral sequences of the DNA are especially suited for the study

of genetic distance because they mutate and evolve in a clock-like manner, providing an elegant way to determine how long ago two species diverged⁶. Data of this kind for Galapagos species are not yet available and are eagerly awaited. If they confirm the protein data on the differences between the finch and iguana lineages, the apparent controversy may disappear. In view of the existence of possible source islands, there is no reason why the birds and the lizards should show identical patterns of

on old islands that are now eroded nearly to sea level and do not now support *Drosophila*. Most of the rest of the lineages form clusters of genetically very close species, meaning that most of them have probably evolved recently, as each newer high island was sequentially colonized from the older ones. The fact that more than 95 per cent of the species are endemic to the high-altitude forests of single islands supports this view. The newest island, Hawaii, is very young (about 0.4 million years) yet harbours many species, including several morphologically and behaviourally differentiated new species found only on this island. Protein comparisons⁸ and DNA-DNA hybridization studies⁹ of one pair of

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Four of the famous finches — Darwin's finches — from the Galapagos. The islands have a special place in the history of evolution. It was observations during his visits to South America and the Galapagos while aboard the *Beagle*, especially of the finch diversity in the Galapagos, that were among the key elements which led to Charles Darwin's eventual formulation of the idea of the origin of species by natural selection.

divergence and descent: one lineage may indeed be new and the other old.

The simultaneous presence of old and new lineages on the Galapagos finds a parallel on the archipelago of Hawaii, another set of drifting islands that has been formed sequentially over a hotspot. The six newer and higher islands, the oldest of which is about 5 million years in age, harbour more than 350 species of *Drosophila* flies found nowhere else in the world. Although these species show great morphological diversity, DNA sequences of the alcohol dehydrogenase gene of selected species show relatively small genetic distances when compared with continental forms⁷. One lineage of fungus-feeding Hawaiian *Drosophila*, however, seems to be about 10 million years old (that is, twice as old as the oldest island where the flies are now found). This case resembles that of the iguanas on the Galapagos, although in Hawaii the lineages may have separated

these species, however, imply that the time of divergence between them must be very recent, not much different from the age of the island.

These multidisciplinary lines of evidence should stimulate biologists to develop new models of speciation for

1. Christie, D. M. *et al.* *Nature* **355**, 246–248 (1992).
2. Cox, A. in *Patterns of Evolution in Galapagos Organisms* (eds Bowman, R. I., Berson, M. & Leviton, A. E.) 11–23 (AAAS, Pacific Division, California Acad. Sci., San Francisco, 1983).
3. Hickman, C. S. & Lipps, J. H. *Science* **227**, 1578–1580 (1985).
4. Polans, N. O. in *Patterns of Evolution in Galapagos Organisms* (eds Bowman, R. I., Berson, M. & Leviton, A. E.) 219–236 (AAAS, Pacific Division, California Acad. Sci., San Francisco, 1983).
5. Wiles, J. S. & Sarich, V. M. in *Patterns of Evolution in Galapagos Organisms* (eds Bowman, R. I., Berson, M. & Leviton, A. E.) 177–186 (AAAS, Pacific Division, California Acad. Sci., San Francisco, 1983).
6. Kimura, M. *J. molec. Evol.* **16**, 111–120 (1980).
7. Thomas, R. H. & Hunt, J. A. *Mol. Biol. Evol.* **86**, 687–702 (1991).
8. Carson, H. L. *Nature* **259**, 395–396 (1976).
9. Hunt, J. A. & Carson, H. L. *Genetics* **104**, 353–364 (1983).
10. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* 203 (Princeton University Press, 1967).

Mary Evans

oceanic archipelagos. The equilibrium theory of island biogeography¹⁰, for example, might be expanded and refined to take into account biotic cycles that accompany both the appearance and disappearance of islands. Oceanic islands continue to provide illuminating perspectives on evolution quite different from those established for continental faunas and floras. □

Hampton L. Carson is in the Department of Genetics and Molecular Biology, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii 96822, USA.

Fresh light on dark ages

John Peacock

It has been a bad twelve months for cosmology. 1991 began with the much publicized demise of the favoured cold-dark-matter model for galaxy formation^{1,2}. Now Uson *et al.*³ have reported the possible discovery of a giant 'pancake' of hydrogen gas in the early Universe — an object that should not exist according to current cosmological theories.

The biggest challenge in cosmology has long been to explain how the Universe evolved from its initial featureless state to its present-day clumped structure of galaxies, clusters and superclusters. We know from observations of the microwave background radiation that the Universe consisted of very uniform ionized plasma at about redshift $z = 1,000$ (since when the Universe has expanded by a factor of $1 + z$). After this, the plasma cooled and recombined so that it no longer scattered photons; cosmology entered an observational 'dark age' from which it only emerged at redshifts close to 5 with the most distant quasars. What happened in between has been the subject of intense speculation, but little solid progress. It is widely believed that, at some stage, excessively dense regions of the Universe started to collapse under their own gravity. But the fundamental debate remains whether this process worked 'bottom-up', with large-scale structure forming through the merging of small sub-galactic units, or 'top-down', with large clusters forming before any stars or galaxies.

The new observational stimulus to this debate comes from astronomers who used the Very Large Array (VLA) radio telescope to detect a large cloud of neutral hydrogen at high redshift³. The discovery (as so often in astronomy) came when they were looking for something else. The telescope was set to study one of the most distant known radio-emitting active galaxies (0902+34 at $z = 3.397$). This was already known from observations of the optical Lyman- α (hydrogen) emission line to be surrounded by an extended cloud of up to 10^{10} solar masses ($M_{\odot}$) of ionized hydrogen⁴. At this redshift, the 21-cm hyperfine transition of neutral hydrogen lies at 323 MHz, which is in one of the

protected radio-astronomy bands at which broadcasting is forbidden; one can therefore look for neutral hydrogen via absorption of the radio emission from the active galaxy. This was found, and in itself provides something of a puzzle. The column density of gas seen is vastly in excess of what would be required to absorb the Lyman- α emission; the neutral gas therefore cannot be a simple foreground screen, and a clumped geometry is required.

Of much greater significance, however, is the fact that neutral hydrogen was also detected in *emission* at a totally separate position about 0.5° away from the radio galaxy. The separation is so large that there is no question of the two objects being related: this is a serendipitous detection of a cloud of neutral gas at high redshift. Strangely, the redshift of the emitting cloud is identical within the errors to that of the absorber in front of the radio galaxy. This may suggest some strange artefact, but Uson and colleagues argue persuasively that this is not the case. Because only a narrow range of redshifts was accessible to observation, the probability of such a coincidence in redshift is about 1 per cent, which is surprising rather than startlingly improbable.

The physics of 21-cm emission is quite simple, so the properties of the cloud can be inferred quite well once some geometry for the Universe is adopted (density $\Omega = 1$ and Hubble expansion rate $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, in what follows). The mass of gas is large, $4 \times 10^{13} M_{\odot}$ (unless it is so clumped that it absorbs some of its own emission, in which case there is even more gas). Better still, the object is resolved both spatially and in velocity: its transverse extent is about 0.1° (about 1 Mpc) and the velocity dispersion is 180 km s^{-1} .

In considering what this object might be, it is helpful to recall the properties of the most massive collapsed structures in the current Universe: clusters of galaxies. Taking as an example the nearby rich cluster Coma, which contains over 1,000 galaxies, the total mass (within a 2 Mpc radius) is about $10^{15} M_{\odot}$, of which about $6 \times 10^{13} M_{\odot}$ is in the form of gas⁵. Given that the volume of space surveyed

(angular extent of map times redshift range covered) is enough to contain about 10 rich clusters like Coma, and given the agreement in gas mass, the obvious interpretation of this object as a high-redshift cluster might seem reasonable. However, there is a big problem: the gas in present clusters is far from being in the form of cold, neutral hydrogen; rather, it is a very hot and highly ionized plasma, which emits X-rays. The neutral content of a typical rich cluster would be less than $10^{12} M_{\odot}$; moreover, this is largely associated with individual galaxies, so that the velocity dispersion would be large (500 km s^{-1}).

Could clusters of galaxies have contained largely neutral gas at early times? To answer this question, we need to say a little about competing pictures for the formation of cosmic structure. As mentioned above, a top-down theory is one in which there is some large coherence length in the primordial mass perturbations; nothing happens until objects on this scale collapse under their own gravity, after which stars and galaxies form through fragmentation. Such a picture is the natural one in a Universe built out of normal matter: dissipative processes in the primordial gas erase small-scale details. This theory was explored in detail during the 1970s by the legendary Soviet school headed by Zel'dovich; their prediction⁶ was that 'pancakes' of neutral hydrogen should form, with masses of about $10^{14} - 10^{15} M_{\odot}$.

Do we therefore now say that this theory is verified, with the first detection of a pancake? Certainly, the object under discussion seems hard to account for in what has been the standard model of galaxy formation for most of the last decade. In the cold-dark-matter (CDM) model, most of the mass is provided by a collisionless, non-dissipative background of exotic non-baryonic particles. Galaxy formation is therefore a bottom-up process of hierarchical mergers between clumps of gas plus dark matter; gas in clusters is kept ionized through continual shock heating caused by new clumps being accreted by a growing proto-cluster⁷. The gas in a pancake is also ionized in its initial collapse, but the difference between this and the CDM model is that the gas in a pancake can subsequently radiate away its energy undisturbed, resulting in the formation of a dense neutral core.

Anyone tempted by the new observations to abandon bottom-up models and return to the old pancake picture faces some severe headaches, as the CDM model was designed to deal with some of the most important developments elsewhere in cosmology. Any tenable theory for galaxy formation must account for the prevalence of dark matter (which exceeds the gas and stars we can see

roughly tenfold in clusters: see above). If this is to be in the form of invisible baryons (for example, rocks), then the successes of the theory of primordial nucleosynthesis⁸ in accounting for light-element abundances tells us that the Universe must be of low density ($\Omega < 0.1$). However, a number of developments recently have seemed to argue very convincingly for a high-density Universe (which must therefore be dominated by non-baryonic material): particularly the high velocity of large-scale galaxy streaming patterns⁹, and the continued absence of fluctuations in the microwave background¹⁰. Perhaps the least radical modification which could account for the new observation is to appeal to dark-matter particles that weigh somewhat less than electrons (rest mass of the order of 1 keV) — 'warm dark matter' (WDM). Such particles would erase small-scale structure though relativistic streaming when they are hot at very early times, leading to a top-down picture once more¹¹. This seems to be the only way of accounting for an apparent pancake while avoiding the great difficulties listed above. Even then, problems will remain: if galaxies cannot predate clusters, then galaxies may need to have formed at embarrassingly low redshifts. However, this is probably a rather minor difficulty compared with trying to construct a baryonic Universe with $\Omega = 1$.

If this observation is confirmed, it will add one more problem to a subject which already seems to be in crisis. Much publicity has been given recently to data on large-scale galaxy clustering which seem to rule out the CDM model¹ (although the difficulties may have been exaggerated²). It may still be that our ideas are fundamentally sound, and that an evolutionary change, such as replacing CDM with WDM, will make everything fall into place. What is more exciting, but also terrifying, is the possibility that such cures will fail: it could mean that almost all cosmological research in the past decade has been in totally the wrong direction. □

John Peacock is at the Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK.

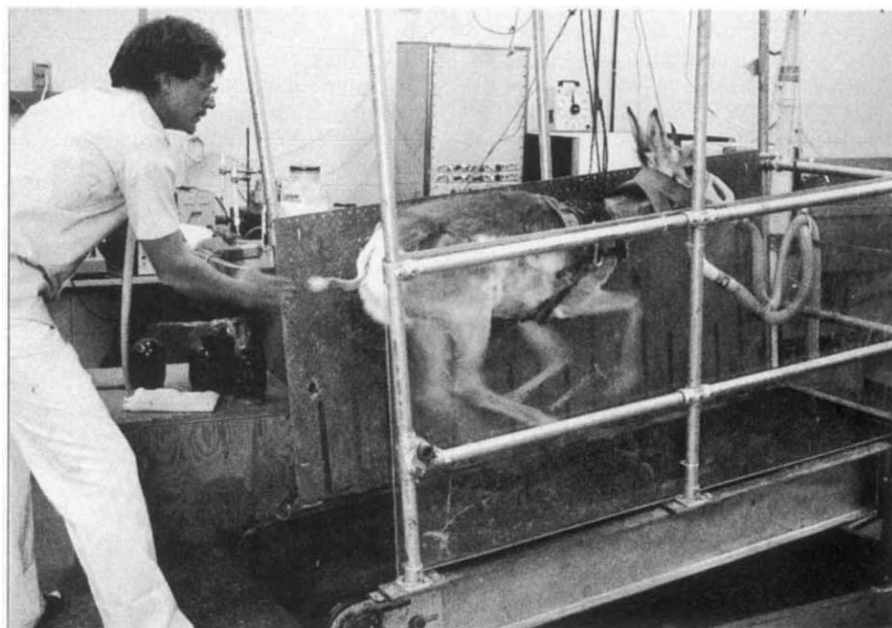
The red flag of optimality

Jared M. Diamond

ONE of Darwin's great achievements was to provide a framework for understanding how an organism's anatomy, biochemistry and physiology are adapted to life tasks. However, most evolutionary discussions of adaptation have been qualitative. How fine is the quantitative match between an organism's capacities and the natural loads upon those capacities? For example, by how much (if at all) do enzyme $V_{\max}$ values exceed natural substrate fluxes, bone strengths

concept of symmorphosis is appealing, it requires testing, because many factors could prevent its realization in practice and because it is still uncertain how closely capacities are matched to natural loads even when symmorphosis does apply.

Weibel, Taylor and Hoppeler¹ have now summarized a decade of testing the applicability of the idea to the mammalian respiratory system. This system is highly suitable for such a test: it serves a



All out effort — a pronghorn antelope on a treadmill is pushed to maximum aerobic capacity (see refs 7 and 8). Photograph by E. R. Weibel.

exceed natural stresses, and intestinal absorptive capacities exceed actual food intakes? Writing in *Proceedings of the National Academy of Sciences*¹, Ewald Weibel, C. Richard Taylor and Hans Hoppeler describe their tests of a promising approach to such questions.

If engineers were designing optimal animals from scratch, they might design for economy: capacities would be matched to peak natural loads, allowing for some safety margin of reserve capacity. As a corollary, series capacities (for example of consecutive enzymes in a metabolic pathway) would be matched to each other. Weibel and Taylor² earlier termed this design principle 'symmorphosis' and reasoned that natural selection would actually tend towards this outcome. Because animals have finite biosynthetic energy and space, animals with large unused reserve capacities squander energy and space that could otherwise be devoted to capacities in short supply, and they tend in fact to be outcompeted by animals of more economical design³. But although the

single dominant function, oxygen delivery from air to mitochondria; it transfers oxygen through a clear sequence of structures with measurable capacities (lungs, heart, capillaries, mitochondria); it performs up to a well-defined natural ceiling, the so-called maximum aerobic capacity (maximum oxygen uptake during peak short-term exercise), with which actual capacity at each step can be compared; and over 90 per cent of that peak oxygen uptake goes to skeletal muscle during all-out quadrupedal running.

Weibel *et al.*'s strategy was to study mammalian species spanning a 1,000-fold range in aerobic capacity, attained partly by selecting species ranging in size from a 20-g mouse to a 450-kg steer, and partly by comparing more and less athletic species of similar size (for example vigorous little dogs and big horses with lazy little goats and big steers). The authors ran each species on a treadmill to measure its aerobic capacity, and then made physiological and morphometric measurements of parameters determin-

1. Saunders, W. *et al.* *Nature* **349**, 32–38 (1991).
2. Peacock, J. A. *Mon. Not. R. astr. Soc.* **253**, 1P–6P (1991).
3. Uson, J. M., Bagri, D. S. & Cornwell, T. J. *Phys. Rev. Lett.* **67**, 3328–3331 (1991).
4. Lilly, S. J. *Astrophys. J.* **333**, 161–167 (1988).
5. Hughes, J. *Astrophys. J.* **337**, 21–33 (1989).
6. Sunyaev, R. A. & Zeldovich, Ya. B. *Mon. Not. R. astr. Soc.* **171**, 375–379 (1975).
7. Evrard, A. E. *Astrophys. J.* **363**, 349–366 (1990).
8. Bøsegard, A. & Steigman, G. A. *Rev. Astr. Astrophys.* **23**, 319–378 (1985).
9. Kaiser, N. *et al.* *Mon. Not. R. astr. Soc.* **252**, 1–12 (1991).
10. Bond, J. R., Efstathiou, G., Lubin, P. M. & Meinhold, P. R. *Phys. Rev. Lett.* **66**, 2179–2182 (1991).
11. Kolb, R. W. & Turner, E. L. *The Early Universe* (Addison-Wesley, Reading, Massachusetts, 1990).

ing capacity for oxygen transfer at each step.

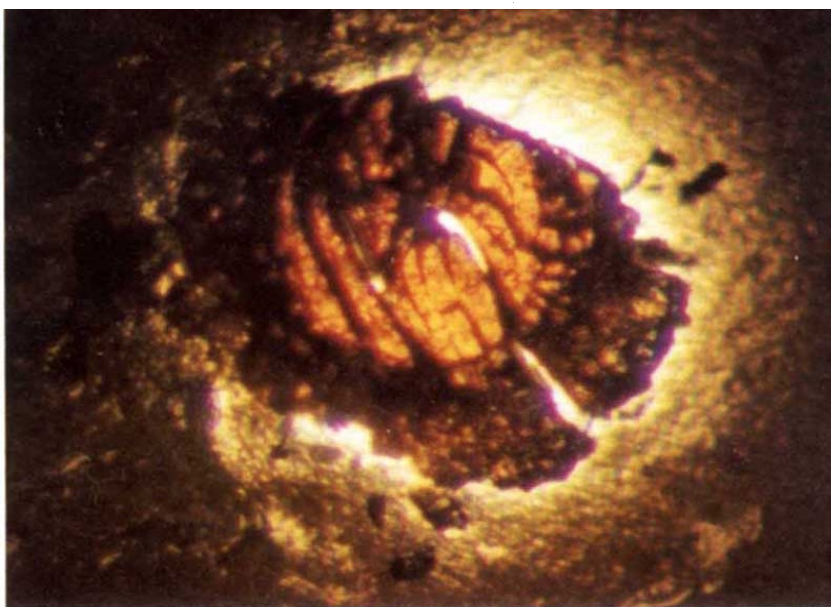
The maximum capacities of several steps could be determined for comparison with actual aerobic capacity. For example, the final step in the mammalian respiratory system is oxygen uptake by mitochondria. Experiments *in vitro*⁴ showed that 1 ml of muscle mitochondria has a maximal oxygen consumption of about 6 ml O₂ min⁻¹ with physiological substrates such as succinate. From actual aerobic capacities and total mitochondrial volumes of each species, Weibel and colleagues calculated that in peak aerobic exercise virtually all skeletal muscle mitochondria of individuals of each species are operating at almost full capacity. That is, there is a close match between theoretical mitochondrial capacity and actual peak load, with little reserve capacity left. In contrast, for the oxygen-diffusing capacity of the lung some reserve remains unused even during a sprint, ranging from 50 per cent of total capacity for a big languid steer to only 20 per cent for a small athletic dog. The heart's ability to deliver oxygen similarly exceeds aerobic capacity by a modest reserve^{3,6}.

Detailed physiological and morphometric studies of individual steps reveal how each step's capacity is adjusted interspecifically to accommodate the 1,000-fold range in aerobic capacity among the mammalian species studied. The simplest case is for oxygen uptake by muscle mitochondria, because there is no evidence of species differences in individual mitochondria. Instead, species with greater aerobic capacities simply have proportionately greater total mitochondrial volumes (for instance 40 times greater in a big swift horse than in a small slow goat; remarkably, Weibel *et al.* actually devised sampling procedures for estimating total mitochondrial volume in these animals). All those mitochondria consume oxygen at a rate of up to 5 ml O₂ min⁻¹ for each millilitre of mitochondria.

The next simplest case is for oxygen transfer across muscle capillaries. All species end up with 1.8 ml of erythrocytes in their muscle capillaries for each ml O₂ min⁻¹ of aerobic capacity, but species with bigger aerobic capacities achieve higher capillary erythrocyte volume in two different ways — smaller species just by having a proportionately greater relative capillary volume, more athletic species partly in that way and partly by possessing a higher haematocrit (erythrocytes constitute a greater fraction of their blood volume).

So muscle mitochondria achieve greater capacity by increasing one factor alone, muscle capillaries by increasing two. Variation in three factors is involved in matching the oxygen content

Squeezing diamond from fullerenes



It may not exactly look like a girl's best friend, but the diamond in the picture above has been formed as never before. Before being squeezed to pressures above 20 gigapascals (GPa) by Núñez Regueiro *et al.*¹, the sample consisted of the black powder that is solid C₆₀.

Interesting consequences have been predicted from compressing C₆₀ crystals since they became available in macroscopic quantities. Perhaps they would become harder than diamonds². Maybe they would link up to form an open, zeolite-type network³. Duclos *et al.*⁴ showed evidence for a change of phase around 16 GPa under non-hydrostatic (anisotropic) compression. The possibility of a pressure-induced transition to a metallic phase motivated earlier experiments by Núñez Regueiro and colleagues⁵; but instead they found a new, insulating phase at 15–20 GPa.

Little could be said then about the nature of this phase, although the authors assumed that theirs was the same as that of Duclos *et al.*, which had been shown by X-ray diffraction to be of low symmetry, and that it contained an appreciable degree of sp³ (saturated) bonding between carbon atoms. Were it not for the potential application of their

results to diamond synthesis, the realization of Núñez Regueiro *et al.* that their present high-pressure phase¹ is familiar old diamond might disappoint those hoping for an entirely new form of solid carbon. But there is no guarantee that the new phases seen in the earlier experiments^{4,5} were diamond too.

The formation of diamond by squeezing is not unprecedented — graphite too can do the rearranging act, but less readily. Its phase boundary lies at a mere 1.6 GPa at room temperature, but because of the very different types of bonding in the two materials, graphite will not make the transition even if compressed far beyond this. Only when shocked to 30–50 GPa or heated above 1,200 K in the presence of a catalyst will graphite convert to diamond. But unlike graphite, C₆₀ contains a degree of sp³ as well as sp² (unsaturated) hybridization, fullerene molecules being non-planar, and this presumably contributes to the ease of the structural transformation at room temperature¹. **Phillip Ball**

1. Núñez Regueiro, M. *et al.* *Nature* **355**, 237 (1992).
2. Ruoff, R. S. & Ruoff, A. L. *Nature* **350**, 663–664 (1991).
3. O'Keeffe, M. *Nature* **352**, 674 (1991).
4. Duclos, S. J. *et al.* *Nature* **351**, 380 (1991).
5. Núñez Regueiro, M. *et al.* *Nature* **354**, 289 (1991).

of blood pumped by the heart to aerobic capacity. Hearts of athletic species pump more oxygen per minute, partly because their blood has a higher haematocrit, partly because the heart itself is larger and pumps more blood per stroke. Hearts of larger species pump more oxygen per minute, partly because the heart is, of course, larger, but the observed relative constancy of heart size (about 0.58 per cent of body mass for large or small mammals) would leave small mammals undersupplied with oxy-

gen, because the metabolic rate for each gram of tissue is greater for smaller animals. The necessary accommodation of small mammals depends entirely on their higher heartbeat frequency. Combining the effects of these three factors, the hearts of all mammals — regardless of size or athletic prowess — end up pumping about 2.8 ml erythrocytes min⁻¹ for each ml O₂ min⁻¹ of aerobic capacity.

These examples make clear that tests of symmorphosis based on comparative

physiology cannot compare single parameters such as haematocrit, but must compare all relevant functional parameters. The reason is that differences in overall performance in several of the above cases arise from differences in combinations of two or three parameters, not from differences in one parameter alone. For example, pronghorn antelope achieve their record-breaking high aerobic capacity by means of large lungs and high haematocrit and high cardiac output and high mitochondrial volume^{7,8}.

Many biologists react instinctively to optimality models as do bulls to red flags. Symmorphosis has attracted vigorous criticism, much of it involving enumerations of reasons why animals' bodies may prove to deviate from optimal economic design^{9,10}. For example, evolutionary achievement of symmorphosis may be thwarted by developmental constraints, design constraints, the need to provide for some reserve capacity as a safety margin, differences in that optimal safety margin depending on cost/benefit trade-offs, and roles of the same capacity in several different pathways imposing different loads.

Each of these factors may indeed mean that symmorphosis is not attained in many cases, but to reject the concept on that basis is to misunderstand its usefulness. Like other optimality criteria, symmorphosis is not proposed in the belief that animals' bodies will prove economically designed throughout. Instead, it is proposed as a null hypothesis of biological design in order to detect when these simple economy-based expectations break down, and thus to be able to analyse the reasons for the breakdown. Without such a null hypothesis, one would not even know that there was a miseconomy of design to be explained. In this sense, symmorphosis seems to me to be a considerable analytical advance that will permit biochemists, physiologists and anatomists to work at the interface between evolutionary biology and cell and molecular biology, and to understand biological adaptation in quantitative terms. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

The evolution of forgiveness

H. C. J. Godfray

VAMPIRE bats cannot guarantee finding a meal every night, but on nights when they do not find food they are often fed by successful bats in the same colony¹. If the two bats are related then the theory of kin selection can account for the evolution of altruistic behaviour such as this. The evolution of altruism among unrelated individuals is, however, much harder to explain. Trivers² argued that natural selection may favour altruism among non-relatives if the two participants meet on a number of occasions so that reciprocal altruism may occur, an idea which meets with the difficulty of understanding how reciprocal altruism can spread in a nonaltruistic population. On page 250 of this issue³, Nowak and Sigmund show that some of the best strategies of reciprocal altruism require the presence of other 'catalyser' strategies in the population before they are able to increase in frequency.

Most theoretical discussion of reciprocal altruism has focused on a particularly simple game called the iterated prisoner's dilemma (the name derives from its other use as a model of human behaviour). This game assumes that a pair of animals meet on a number of occasions and at each meeting behave in one of two ways: cooperate or defect (act selfishly). The fitness of each animal is influenced by its own behaviour, and the behaviour of the participant. An individual does best when it defects and its partner cooperates (most commonly the payoff to the defector is taken as five units while the suckered cooperator receives zero). If both individuals cooperate, they each receive a moderate payoff (three units) while they both obtain a low payoff (one unit) in cases of double defection. Although mutual cooperation appears to be the best strategy, the advantages of defecting when your partner cooperates renders it susceptible to cheating. If the two animals meet only once, or just a few times, the optimal strategy is always to defect. But when they meet on a number of occasions, reciprocal altruism becomes a possibility. The problem is to discover the best strategy based on the animal's previous experience with the other individual.

The optimal solution to the iterated prisoner's dilemma is particularly difficult to obtain because of the huge number of possible strategies. A novel approach was adopted by Axelrod⁴, who invited biologists and game-theorists to send in strategies whose performances in competition with each other were compared in a computer tournament. Rather surprisingly, the winner was the simple

strategy: on the first encounter cooperate; on subsequent encounters do whatever the other animal did on the previous encounter. This do-as-you-would-be-done-by strategy is called tit-for-tat. Once established, tit-for-tat cannot be invaded by any alternative strategy, although many others perform as well as tit-for-tat.

Although tit-for-tat did best in the computer tournament, the set of rules may not have been representative of the possible behavioural strategies used by real animals. In addition, stochasticities such as errors of recognition, likely to be important in nature, were given little attention⁵. The likelihood of such random events has now led Nowak and Sigmund to investigate a series of stochastic strategies related to tit-for-tat (see also refs 6–8). Here, the animal always responds to past cooperation by itself cooperating, but instead of always responding to defection by further defection, it forgives defection on a proportion (q) of occasions. It can be shown that for the usual parameters of the prisoner's dilemma, a population adopting this strategy (called GTFT for 'generous tit-for-tat') with $q = 1/3$ cannot be invaded by strategies with different probabilities of forgiveness.

Although GTFT may resist invasion by other strategies, can it itself invade populations playing by different rules? This was investigated by running computer tournaments with a wide range of strategies including GTFT and the strategy 'always defect'. In tournaments where tit-for-tat was not included, the strategy of always defecting normally won. In the initial stages of the competition, defection is a good policy as there are many individuals in the population willing to cooperate who can be exploited for large payoffs. Once the population is largely composed of defectors, there is no advantage to forgiveness and GTFT does badly. However, if tit-for-tat is included in the tournament an interesting result occurs. First, the always defect strategy increases in fitness for the same reasons as before. When defectors make up most of the popula-

1. Weibel, E., Taylor, C. R. & Hoppeler, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10357–10361 (1991).
2. Weibel, E. & Taylor, C. R. *Respir. Physiol.* **44**, 1–164 (1981).
3. Dykhuizen, D. *Evolution* **32**, 125–150 (1978).
4. Schwarzmann, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1583–1587 (1989).
5. Taylor, C. R. et al. *Respir. Physiol.* **69**, 7–26 (1987).
6. Jones, J. H. et al. *J. appl. Physiol.* **67**, 862–870 (1989).
7. Alexander, R. M. *Nature* **353**, 696 (1991).
8. Lindstedt, S. L. et al. *Nature* **353**, 748–750 (1991).
9. Garland, T. & Huey, R. *Evolution* **41**, 1404–1409 (1987).
10. Dudley, R. & Gans, C. *Physiol. Zool.* **64**, 627–637 (1991).

1. Wilkinson, G. S. *Nature* **308**, 181–184 (1984).
2. Trivers, R. L. *Q. Rev. Biol.* **46**, 35–57 (1971).
3. Nowak, M. A. & Sigmund, K. *Nature* **355**, 250–253 (1992).
4. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
5. May, R. M. *Nature* **327**, 15–17 (1987).
6. Nowak, M. & Sigmund, K. *J. theor. Biol.* **137**, 21–26 (1989).
7. Nowak, M. & Sigmund, K. *Acta appl. Math.* **20**, 247–265 (1990).
8. Nowak, M. *Theor. Pop. Biol.* **38**, 98–112 (1990).
9. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).

tion, tit-for-tat increases in frequency. Individuals playing tit-for-tat respond aggressively to obligate defection, yet cooperate among themselves. Finally, once tit-for-tat has eliminated most of the obligate defectors, GTFT is able to invade. Thus GTFT is the eventual winner, but only thanks to the midwifery of tit-for-tat.

What do such abstract games as the iterated prisoner's dilemma tell us about reciprocal altruism in nature? First, they suggest that natural selection may favour relatively simple rules of conduct, a comforting thought for experimentalists. Nowak and Sigmund's work further implies that these rules are likely to be probabilistic, and more forgiving than DNA PROFILING

simple tit-for-tat. It is less clear whether the type of evolutionary dynamics observed by Nowak and Sigmund in their computer experiments will also be seen in the field. It may be far easier for strategies of reciprocity to invade in nature than in computer tournaments if, at least initially, there is a high probability of interactions between relatives⁹. So perhaps we need to consider not only the prisoner's dilemma, but also the dilemmas confronting the prisoner's relatives. □

H. C. J. Godfray is in the Department of Biology and the NERC Centre for Population Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK.

Law and probabilities

John Brookfield

LAST month, in *Science*¹, the population geneticists R. C. Lewontin and D. L. Hartl laid down a challenge to the way in which the probabilities associated with DNA profiles are presented in courts; DNA profiling is used routinely in forensic science in Britain and the United States, and can provide powerful evidence against defendants. Lewontin and Hartl say that current methods of calculation produce serious errors, which may result in false incriminations, a claim that has evoked controversy^{2,3} and has received a sharp rejoinder from R. Chakraborty and K. K. Kidd⁴.

The usual use of DNA profiles in courts is in showing that, at a series of 'variable number of tandem repeat' (VNTR) loci, the DNA of a suspect matches DNA obtained from the scene of a crime. A match constitutes evidence against the suspect, the strength of the evidence depending on the probability that an innocent person would produce a match. It is with the calculation of this probability that Lewontin and Hartl are concerned.

The standard practice is to calculate the probability of a match at a series of loci as the product of the probabilities of matches at each of the loci individually. For this product to represent the true probability, the probabilities of matching at the individual loci must be independent of each other and thus there must be linkage equilibrium between these loci in the population. If there is linkage disequilibrium, in the form of racial or subracial structuring, then the multiplication of probabilities will give inaccurate values, which will generally be too low. For example, if two individuals match at four VNTR loci, the matching at the first one or two will provide evidence that the two individuals are

from the same population subgroup. This, in turn, will mean that their probabilities of matching at the subsequent loci will be increased relative to random individuals.

The problem is compounded if it is known, in advance, that the truly guilty person and the suspect are from the same ethnic subgroup. There is plenty of evidence of differences in the frequencies of VNTR alleles between three of the main racial groups in the United States (caucasians, blacks and hispanics) and, generally, in courts the probability given is that corresponding to the group which maximizes this probability. But Lewontin and Hartl say (rightly) that these three broad categories are themselves heterogeneous collections, and could have further structure within them. Such structuring has been found for blood group loci, distinguishing, for example, Poles and Italians. Probabilities calculated using databases compiled from heterogeneous collections of caucasians may, they argue, underestimate the true probabilities of matches by two orders of magnitude.

No one disagrees that probability estimates based on the multiplication of probabilities from individual loci require linkage equilibrium between the loci, and that the more linkage disequilibrium there is, the more inaccurate they will be. But in their response⁴, Chakraborty and Kidd point out that the amounts of linkage disequilibrium will be sufficiently small that the inaccuracies introduced in probability calculations will be very minor. They show that levels of disequilibrium known for blood group data, although considerable, would not induce serious quantitative inaccuracies in probabilities derived using multiplication rules. The same will, they argue, prob-

ably be true for the subpopulation structuring of VNTR loci, and indeed what data we now have strongly support this view.

Chakraborty and Kidd's examples show that quite large amounts of divergence in allele frequencies between subpopulations produce only small inaccuracies in probability calculations if the appropriate database is used. For estimates to be seriously inaccurate, populations would have to consist of subpopulations, each characterized by a small number of common alleles, each of which is rare in the population as a whole. Surveys of diverse populations using single-locus⁵ or multilocus DNA profiling⁶ would have been expected to have revealed this pattern if it existed, but they have not.

Lewontin and Hartl contend that there is no way of guessing how large VNTR substructuring will turn out to be, and thus no way of knowing the inaccuracy of probability estimates assuming none. They propose three methods for giving estimates which will not falsely incriminate any suspect. First, that the probability of a chance match should be expressed simply as the proportion of times that multilocus genotype has been observed in the appropriate database; this approach means that the minimum probability of a chance match that can be given is the reciprocal of the number of individuals in the database. The other methods would include use of details of allelic frequencies in all subpopulations, and the production of probabilities using all this information, or, if it is partially lacking, the use of the maximum frequency for each allele that has been observed in any subpopulation. Lewontin and Hartl's conclusion, and the most contentious of their recommendations, is that DNA evidence should be regarded as unreliable unless and until we have the detailed information on population structure that will make these new approaches possible.

The recommendation, if stringently applied, would be a setback for forensic science. It would prevent DNA profiling from being used to incriminate any suspect, for however many years it will take to gather the information concerning population structuring. This would include cases in which there was DNA evidence that constituted a strong indictment of a suspect under almost any conceivable degree of subpopulation structuring, and in which that evidence formed the bulk of the prosecution's case. Crucially, a statement of a probability is, by its nature, a statement of partial knowledge, so it is paradoxical to imply that in principle we cannot calculate the probability of an event without further empirical knowledge. Given present knowledge of the degree of

population substructuring for other loci, it is possible, in principle, to predict the likelihoods of various degrees of population substructuring at VNTR loci, to calculate the probabilities that these substructuring patterns will cause a match to an innocent person, and thereby, by the use of a weighted average, to calculate the true probability of a match in any particular case. In practice, however, we will not be able to calculate this true probability, and an estimate has to be adopted. The multiplication rule gives one estimate. Any bias downwards in this estimate as a result of population substructure will be small in almost all cases compared to the sources of conservativeness that are built into the procedure at other points.

Lewontin and Hartl's proposals of new methods have the advantage that the methods are designed to be sufficiently conservative that the probabilities that they calculate will always be equal to or greater than this true probability. This means that discrepancies between the probability quoted in court and the true probability will never lead to false convictions. There are still snags, however. First, although the argument is that they should be used only when we have much greater knowledge of population substructuring, there can be no finite population survey which could demonstrate absolutely that these methods will always be conservative. Second, although such methods consistently produce estimates of probability biased in favour of the suspect, there will be no consistency across cases in the size of this bias, which will probably be affected strongly by the racial structure of the community in which the crime was committed.

The animosities being generated in the current round of exchanges are unhelpful, and indeed surprising given the extent of common ground shared by the protagonists. The human danger implicit in what to many will seem esoteric scientific arguments is the creation of a situation of unreasonable doubt, in which the guilt or innocence of individual suspects is determined in haphazard fashion by the willingness of individual courts to accept DNA evidence as admissible. □

John Brookfield is in the Department of Genetics, Queens Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK, and is a consultant for Cellmark.

Synthetic gels on the move

Kanji Kajiwaru and Simon B. Ross-Murphy

CERTAIN synthetic polymer gels can be said to be intelligent, that is they can be made to respond to changes in their environment — pH, temperature, voltage and so on — by changing their shape, solubility or their degree of ionization, for example. Such gels are much more than mere curiosities — two meetings have been devoted to them in the past two years; a new journal (*Journal of Intelligent Material Systems and*

trollable membrane separation systems and electronically regulated drug delivery systems. A sophisticated application of such intelligent gels for the second purpose was reported by I. C. Kwon *et al.* only very recently (*Nature* **354**, 291–293; 1991). In their proposed system a mixed poly(acrylic acid)/poly(ethyl-oxazoline) gel can be dissolved when a current is passed through it, and the authors propose to make this the basis of a long-period, controlled-release device that could, for example, furnish insulin for diabetics.

Hydrogels prepared from polymers which show lower critical-solution-temperature behaviour in aqueous solutions are thermoresponsive, and expand or contract below or above the phase transition temperature. The detailed application of such materials is based upon the studies of T. Tanaka and co-workers at the Massachusetts Institute of Technology over the past dozen years. A gel containing ionic groups can be actuated isothermally by an electric field. When the gel is negatively charged, it



Gripping stuff: the newest polymer-gel fingers grasp a piece of paper at the recent RoboBug Fest.

Structures, Technomic Publishing) has been launched to carry news of developments; and, elsewhere in this issue (*Nature* **355**, 242–244; 1992), Y. Osada and colleagues describe how one such gel can be invested with 'worm-like' motility.

A thermodynamic system capable of transforming chemical energy directly into mechanical work is known as a chemomechanical system. The history of the materials involved goes back to the early 1950s and such distinguished names in polymer science as W. Kuhn and A. Katchalsky, who described a number of aspects of the chemomechanical response. All living organisms move by the isothermal conversion of chemical energy into mechanical work as exemplified by muscular, flagellar and ciliary movement. Such highly efficient energy conversion systems can now be realized in synthetic polymer gels, which expand and contract upon changing their solubility and/or degree of ionization according to an external stimulus supplied in the form of thermal, chemical and electrical energy. Many artificial chemomechanical systems made of environmentally sensitive polymer gels are being developed for sensor/actuator systems such as con-

swells near the anode and contracts near the cathode, the contraction rate being proportional to the external electrical current.

The absolute change in volume is by no means insignificant — dimensional changes of say 50 per cent are quite usual. Such a gel bar made of polyelectrolytic material can bend backwards and forwards by the application of an alternating electrical field. Here, water and ions migrate towards the electrode bearing a charge opposite in sign to the net charge in the gel, and this coupling of electroosmosis and electrophoresis is thought to be responsible for the observed chemomechanical behaviour. A gel bar prepared from poly(vinyl-alcohol) and poly(acrylic acid) — PVA-PAA — bends towards the anode under a static electric field. A robot hand with four fingers of PVA-PAA gel can be made to grasp a fragile egg in an aqueous Na_2CO_3 solution under an applied field of 50 V. And an artificial fish has been made that swims at a speed of 2 cm s^{-1} by waving its 'tail' as the polarity of the field is alternated. In their paper in this issue, Osada and co-workers describe in detail the behaviour of a hydrogel that can move through a bath of

1. Lewontin, R. C. & Hartl, D. L. *Science* **254**, 1745–1750 (1991).
2. Anderson, C. *Nature* **354**, 500 (1991).
3. Roberts, L. *Science* **254**, 1721–1723 (1991).
4. Chakraborty, R. & Kidd, K. K. *Science* **254**, 1735–1739 (1991).
5. Budowle, B. *et al.* *Am. J. hum. Genet.* **48**, 841–855 (1991).
6. Jeffreys, A. J., Turner, M. & Debenham, P. *Am. J. hum. Genet.* **48**, 824–840 (1991).

RÉSUMÉ

electrolyte solution. The system consists of a poly(2-acryl-amido-2-methyl propane sulphonic acid) gel doped with *n*-dodecyl pyridinium chloride surfactant. They show how a ratchet-type 'gel looper' can 'crawl' along a bar when a 20 V voltage is supplied.

At the most recent meeting of the Society of Polymer Science of Japan Polymer Gels Group, held last month in the science city of Tsukuba, a number of demonstrations of intelligent gel technology were shown at the so-called 'Robo-Bug Fest'. Examples included a polymer gel finger that works in air (see figure) rather than in solution, a significant advance; a gel beetle that can crawl up an incline; a gel-powered rotational micromotor; a flapping gel wing; and a gel Ferris wheel. A notable aspect of the progress in this area is the powerful synergy of university, government and industrial research, a liaison driven by the Agency of Industrial Sci-

ence and Technology of the Japanese Ministry of International Trade and Industry.

At the moment the study of externally responsive gels is still in its infancy, and there remain some potential problems before they can be successfully exploited in terms of applications. For example, the gels respond rather slowly, typically over timescales of minutes. But the time required for a spherical gel to shrink is proportional to the square of its radius, so that sufficient speed could be obtained by miniaturization, and the construction of gel microfibrils is already being discussed. Applications might then include the construction of devices which can be inserted into muscle to emulate a nerve response. □

Kanji Kajiura is at the Kyoto Institute of Technology, Kyoto, Sakyo-ku, Matsugasaki, 606 Japan. Simon B. Ross-Murphy is in the Biomolecular Sciences Division, King's College London, London W8 7AH, UK.

Lost and found

To lose a type specimen is unfortunate, but to lose an entire collection might be construed as carelessness — especially if part of it turns up after nearly 200 years under one's very nose. It also causes confusion, as happened after the insect collections of the Swedish entomologist G. J. Billberg were destroyed by fire in 1822. The collections contained many type specimens of important tropical beetle species, and their loss has long inhibited beetle taxonomy. But it has turned out that Billberg's brother-in-law, C. J. Schönherr, was also an entomologist, and that the two had traded specimens. As J. Ferrer discovered (*Annals of the Transvaal Museum* **35**, 279–283; 1991), seven of Billberg's lost type specimens were indeed buried, unrecognized, in Schönherr's own collections, kept at the Naturhistoriska Riksmuseet in Stockholm.

Martian moraines

VAST ice sheets once lay over the poles of Mars, covering as much as one-fifth of the planet's surface, say J. S. Kargel and R. G. Strom (*Geology* **20**, 3–7; 1992). The authors compared Viking pictures of the planet's surface with glacial landforms on Earth. Extensive ridge patterns, they argue, can only be eskers, sand-and-gravel deposits laid down in channels below ice sheets. Drumlins and subglacially eroded tunnel valleys are also identified. From the cratering record, Kargel and Strom suggest that the glaciation (which lasted perhaps 20 million years) may have occurred as little as 250 million years ago. If correct, their interpretation supports the notion that the planet has known periods much warmer than its present CO₂-freezing climate.

Shooting match

THE beginning of the end of the dishonourable history of the Dum-Dum bullet may be marked by a report by R. M. Coupland *et al.* in *The Lancet* (**339**, 35–37; 1992). Dum-Dums and other bullets which fragment easily after hitting their targets, and thereby cause particularly nasty wounds, were outlawed by the Hague Declaration of 1899. But until now there has been no objective way of knowing if such ammunition has been used. Coupland *et al.* looked at 1,287 wound radiographs from four hospitals and show that the extent of bullet break-up and wound severity can be assessed from measurement of degree of bone fracture and of metallic fragments on the radiograph. They were able to identify one hospital with an especially high incidence of wounds caused by bullet fragmentation, and point out that their study has important implications for upholding humanitarian law.

CANCER

Carcinogens leave fingerprints

Bert Vogelstein and Kenneth W. Kinzler

DETECTIVES on the trail of the causes and mechanisms of cancer have been working hard to connect specific environmental agents with particular molecular changes that occur in carcinogenesis. This pursuit has now led^{1,2} to the indictment of two suspects who left their fingerprints at the scene of the crime. The suspects, aflatoxin B₁ and ultraviolet light, have long been on the wanted list because of their suspected involvement in liver and skin tumours, respectively.

The new findings are the result of the convergence of two areas of inquiry. First, epidemiology has indicated that certain environmental factors are associated with an increased incidence of cancer; these associations cannot by themselves illuminate mechanism, but because many of the factors concerned are genotoxic it has seemed that induced mutations might be to blame. Second, molecular genetics has identified genes that appear to be responsible for the initiation and/or progression of tumours. Are these oncogenes and tumour suppressor genes the primary targets of environmental factors?

Because oncogenes were identified before tumour suppressor genes, they were the first to be examined in this regard. It has now been exhaustively demonstrated that rodent cancers induced by exposure to a variety of carcinogens often contain mutations in *ras* oncogenes (see any recent issue of the journal *Molecular Carcinogenesis*). Importantly, the base changes in the mutant *ras* genes are

generally characteristic of those caused by the carcinogens in simple mutagen assays³, implying that the carcinogen can interact directly with mammalian genes and produce mutations at the site of interaction. A cell which thereby acquires a mutation in a *ras* gene is endowed with a selective growth advantage and clonally expands to form a tumour.

Similar studies on *ras* genes in human tumours have not been as productive, for two reasons. First, a relatively small number of human tumour types contain *ras* mutations at significant frequency⁴. Second, and more important, there are only a few codons in *ras* which, when mutated, result in oncogenic activation *in vivo*. These facts limit the comparisons that can be made and the conclusions that can be deduced from the examination of mutations in different tumours.

In contrast, some suppressor genes (in particular, p53) are mutated in a significant fraction of many different types of human cancer. Moreover, such genes can be inactivated by mutations at numerous sites, through point mutations, deletions or insertions. The first suggested link between a specific carcinogen and p53 came last year when Harris, Ozturk and their colleagues reported that G→T transversions at codon 249 (arginine to serine) were found in hepatocellular carcinomas (HCC) from southern Africa and east Asia; mutations at this position were rarely found in tumours of the colon, lung, breast or

other organs^{5,6}. The data suggested that one of the risk factors endemic to these areas, aflatoxin B₁ or hepatitis B virus (HBV), was responsible for the mutations. However, it was possible that the mutations were not mutagen-specific but rather were tumour-type specific, so that HCC from regions without these risk factors might also contain the same mutations.

This alternative explanation is made less likely by the new data from Ozturk and collaborators¹, which show that HCC from low-aflatoxin exposure areas (even those with a high HBV infection rate) only rarely contain the 249^{ser} mutation. Combined with the fact that aflatoxin B₁ is known to induce G→T transversions in simple model systems, the most likely interpretation of the data is that aflatoxin B₁, in the context of chronic viral hepatitis, drives HCC in humans by forming deoxyguanosine adducts at codon 249, resulting in substitutions of T for G.

Another strong case is made by Brash *et al.*² for the involvement of p53 mutation in human squamous cell carcinomas (SCC) of the skin. Over half of the invasive SCC studied had p53 gene mutations, all of the mutations being at dipyrimidine sites (such a predominance is characteristic of UV light-induced mutations in model systems). Moreover, a quarter of the mutations resulted in CC to TT double-base changes. These double-base changes are virtually pathognomonic of a UV-based aetiology and have rarely been observed in internal tumours.

The analysis of p53 in other tumour types is also proving to be revealing. Lung and oesophageal cancers frequently contain base-pair changes characteristic of those produced by the mutagens present in tobacco smoke and alcohol, respectively^{7,8}. These agents are known to be risk factors from epidemiological studies. In the future, we might expect somatic mutations to point towards specific aetiological factors in cases where epidemiologists are unable to pinpoint specific agents.

Many questions about the relationship between environmental mutagens and cancer remain, even for HCC and SCC. Cancer apparently arises through the sequential accumulation of mutations in several oncogenes and/or suppressor genes⁹. Both HCC and SCC probably contain at least four or five such mutations in each tumour. Are all these mutations induced by aflatoxin B₁ (or sunlight) or do these agents simply place the final straw on the camel's back? This question will become tractable when more mutant genes in these tumours are identified.

The data also have implications for a debate about the nature of human carci-

nogenesis. It has recently been argued (1) that human exposure to industrially derived carcinogens is minuscule compared to that derived from exposure to naturally derived carcinogens, and (2) that carcinogens can act by direct adduct-driven mutagenesis and/or by causing increased cellular turnover (mitogenesis) which indirectly increases mutation rates¹⁰. The new reports bear on these arguments in several ways. Although some people are convinced that technology is responsible for most of society's ills, these first clear examples of carcinogen-induced mutations in human cancer point towards natural rather than man-made agents. Second, the base-pair changes in SCC and HCC p53 genes strongly suggest that they are the result of adduct-driven mutagenesis, rather than secondary effects of increased mitogenesis (although the results cannot rule out another role for carcinogen-induced mitogenesis in the development of these tumours). Third, one might ask why, if the dose of naturally derived mutagens (either from normal oxidative metabolism or from chemicals endogenously present in foods) is so high, is the additional exposure to *single* mutagens such as aflatoxin B₁ and sunlight rate-limiting for cancer development?

This new molecular epidemiology also provides lessons for cancer prevention. It is one thing to tell a tanning champion from California (or a regulatory official in Asia) that epidemiological and animal evidence suggest certain preventative measures. It is quite another to say we know that light of 300 nm induces mutations at pyrimidine dimers in skin-cell p53 genes which will prevent that gene from suppressing tumour growth. The latter is more compelling because of its clarity and inherent logic, providing a coherent picture of how a specific agent contributes to the development of human cancer in molecular detail. The potential economic and human benefits of putting this new knowledge into practice, for prevention of cancer, are immense. □

Bert Vogelstein and Kenneth W. Kinzler are at the Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, Maryland 21231, USA.

1. Ozturk, M. and collaborators *Lancet* **338**, 1356–1359 (1991).
2. Brash, D. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 10124–10128 (1991).
3. Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1987).
4. Bos, J. L. *Cancer Res.* **49**, 4682–4689 (1989).
5. Hsu, I. C. *et al. Nature* **350**, 427–428 (1991).
6. Bressan, B., Kew, M., Wands, J. & Ozturk, M. *Nature* **350**, 429–431 (1991).
7. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *Science* **253**, 49–53 (1991).
8. Bennett, W. P. *et al. Oncogene* **6**, 1779–1784 (1991).
9. Fearon, E. R. & Vogelstein, B. *Cell* **61**, 759–767 (1990).
10. Ames, B. N. & Gold, L. S. *Science* **249**, 970–971 (1990).

Noses have ears

THE smaller the animal, the higher the range of sound frequencies that it can generate and hear. Elephants have an upper limit of hearing well below the human one of 15 kHz or so; cats can hear up to 50 kHz and bats up to 120 kHz. All this, says Daedalus, makes good physical sense. Smaller animals have smaller ears and voice boxes, whose smaller resonant structures create and respond to higher frequencies. Extending the argument, Daedalus expects insects to communicate in the 1-MHz band. Modern ultrasonic technology can easily generate and detect frequencies around 1 MHz, even though in air they travel only short distances. Already ultrasonic generators are sold to discourage cockroaches and similar pests. This simplistic technology could easily be developed. By detecting, decoding and imitating the ultrasonic communications of ants, hornets and other social insects, specific messages could be directed at them. They could be directed not to seek food in our buildings, and to make their nests away from them.

Bacteria, much smaller even than insects, should communicate at 1 GHz and above. This sonic region encroaches into the range of molecular rotational and even vibrational frequencies. According to one theory, our sense of smell works by detecting such molecular vibrations. Clearly, says Daedalus, at these high frequencies the distinction between an ear and a nose disappears. A bacterium must carry on its surface a range of sensory molecules or molecular clusters which respond to specific molecules nearby, either by direct vibrational contact, or by detecting sonically radiated molecular vibrational modes.

So Daedalus is devising non-contact toxins to kill or control bacteria. They never even touch the target micro-organism, but merely radiate the appropriate molecular frequencies. Such a toxin might, for example, imitate the vibrational 'smell' of a nutritional molecule. The organism would open its cell wall in anticipation, and leak to death. Daedalus is vacuum-coating fine particles of bacterial poisons like phenol and penicillin, with very thin layers of gold, magnesium fluoride and so on. Such coatings have few vibrational modes of their own, but should faithfully transmit those of the molecules within. If these coated particles turn out to damage or influence bacteria, the way will be open for a whole new range of micro-encapsulated drugs. Ferociously reactive, they will withstand body chemistry and exert no damaging side effects, because they will never escape from their vibro-transparent envelopes. David Jones

Database contamination

SIR — We have examined sequence data from vertebrate and bacterial species available in update 27 of the EMBL sequence database, using the program FASTA with the complete M13mp18 sequence as well as its polylinker only as search strings, and found extensive homologies to these to be present in several submitted sequences. We found 78 sequences with a total of 81 occurrences of vector sequences not documented in the feature table. Of these 81 overlaps, 55 were found in the 5' or 3' end of the sequence, while 23 were found internally (and presumably used in the assembly of the sequence) and 3 covered essentially the whole submitted sequence. In 19 of the 35 instances where the vector-containing parts were included in the feature table they were described as part of introns, whereas 7 listed them as part of the coding sequence. The largest overlap reported by FASTA contained 599 nucleotides.

Because FASTA will report only the one area in a sequence with the highest similarity to the search sequence, some of the vector-containing sequences reported were investigated further using WORDSEARCH, which will unravel occurrences of several patches of vector sequence. In one case, a sequence of 2,537 nucleotides where FASTA reported an overlap of 562 nucleotides with M13mp18, WORDSEARCH revealed the presence of four regions (125–412, 691–1,183, 1,538–1,700 and 2,155–2,533) with a total of 1,319 bases that clearly (assuming a rather high frequency of reading mistakes) are M13-derived.

Although M13-based vectors have been the most widely used vectors for DNA sequencing, we assume that searches with different polylinkers, lambda and other vector-based sequences as search strings would give several new overlaps. Also, there are probably entries in the databases with vector components too small to be assigned as such by mere homology considerations. The real number of vector-derived sequences in the databases thus is considerably larger than the number reported here.

The abundance of sequencing errors in these vector sequences is also of interest. In some instances, of course, this may be due to a heavily mutated vector being used, but in most cases we expect these errors to be a result of sloppy sequencing. If this reflects the quality of the total sequence information available in the databases, it must influence the reliability of these data as basis for biologically significant conclusions.

Finally, looking at the date of submission of the sequences, there are no

indications that vector sequence contamination is a thing of the past, as 30 of the sequences we found were submitted in 1990 and 11 in 1991.

How can we avoid errors like these finding their way into the databases? Most sequencing software packages contain some sort of vector screening. Thus a simple screening of raw sequence data before assembly, as well as screening of assembled sequences against vector databases before submission, should be a simple task. Another possibility is that database administrators should screen all submitted sequences against the vector database. But is it part of their job to function in this way? One thing they could do is to add a question to the sequence submission forms: "Has this sequence been checked for the presence of vector sequences?". Then the submitters would at least have been made aware of the possibility of contamination of the data. Also, it probably would be useful if submitters were encouraged to include the vector used in the feature

table.

What should be done to contaminated sequences? First, such sequences constitute a very small percentage of the total database. In many cases, it would be enough to include the presence of vector data in the feature table of the sequence. In several instances, we presume that complete removal of the sequences from the database would be the best line of action. Again, it is understandable if the database administrators are reluctant to perform this cleaning-up: preferably, the submitters themselves should be made aware of the doubtful nature of their submissions and given the opportunity to rectify the data or withdraw their submission. (A list of the contaminated sequences we found is available on request from us, and has been submitted to the EMBL Data Library.)

RODRIGO LOPEZ*

TOM KRISTENSEN

HANS PRYDZ

The Biotechnology Centre of Oslo,
University of Oslo,
P.O. Box 1125,
Blindern N-0316 Oslo, Norway
*The Norwegian EMBnet node

Death by superantigen

SIR — Cohen *et al.*¹, commenting on our report² on superantigen-induced peripheral T-cell tolerance, contended that the T-cell death observed in SEB-primed mice may not reflect *in vivo* activation-triggered cell death but rather the *in vitro* death of SEB-activated T cells consequent to lymphokine "starvation". To support this contention, they cite their detection of DNA fragmentation and subsequent cell lysis in interleukin-2 dependent T-cell clones 12 hours after interleukin-2 withdrawal *in vitro*³.

Cohen *et al.* suggested that in our system, SEB remaining in the circulation continues to stimulate V β ⁺ T cells, such that at day 2 post-priming T-cell activation is still in progress and accordingly lymphokines are sufficiently abundant

that the cells do not undergo DNA fragmentation, even after 3 hours of culture *in vitro*. They postulate that instead, by day 4 post-priming, SEB is cleared from the system and the lack of ongoing SEB stimulation leaves V β ⁺ T cells relatively lymphokine-"starved", a situation that is exacerbated in culture resulting in programmed cell death. Therefore, they ascribe the DNA fragmentation we observed in V β ⁺ T cells at day 4 post-priming to a lack of growth factors necessary for cell survival rather than to activation-induced programmed cell death.

The explanation proposed by Cohen *et al.* is interesting, but we believe that their interpretation of our results is not supported in the context of the following observations: (1) We found² that at 2 days after SEB priming, the percentage of V β ⁺ T cells in the periphery was elevated from 20 to 40% both before and after 20 hours' incubation at 37 °C. If as suggested by Cohen *et al.*, *in vitro* cell "starvation" were the primary cause of cell death, one would have expected to see a reduction in the V β ⁺ T-cell number after 20 hours of culture. (2) Anti-interleukin-2 antibodies added to the 37 °C 20-hour culture of spleen cells obtained at 2 days post SEB-priming to block the possible effect of autocrine interleukin-2 do not induce programmed cell death of V β ⁺ T cells. (3) Splenic cells and serum obtained 36 hours after SEB priming did not induce proliferation of fresh spleen cells. Thus, the absence

Scientific Correspondence

SCIENTIFIC Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Due to space limitations priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language. Contributions may be sent to referees and, in the case of matters arising from material published in *Nature*, are often sent to the author of that article for comment.

Scientific Correspondence submitted for publication should be typed double-spaced and three copies sent. □

of detectable $V\beta 8^+$ T-cell death 2 days after SEB administration cannot be attributed to ongoing SEB-induced cell activation.

In summary, our results do not favour the contention of a key role for T-cell lymphokine starvation *per se* in the induction of cell death after SEB-priming, but indicate that diminution of the $V\beta 8^+$ T-cell population in this context is more satisfactorily explained by activation-induced programmed cell death.

ATSUO OCHI

Division of Molecular Immunology
and Neurobiology,
Samuel Lunenfeld Research Institute,
Mount Sinai Hospital,
600 University Avenue, Toronto,
Ontario M5G 1X5, Canada

YOJIRO KAWABI

The 1st Department of Internal
Medicine,
Nagasaki University School of Medicine,
7-1 Sakamoto machi,
Nagasaki 852,
Japan

COHEN *ET AL.* REPLY — Ochi and Kawabe's response does not support their contention. They say that if lymphokine withdrawal was responsible for $V\beta 8^+$ cell death, they would have seen a reduction in $V\beta 8^+$ cell number after 20 hours' incubation *in vitro*. That is exactly what they did see at day 4 and 7 (ref. 2). That they did not see it at day 2 supports our suggestion that at that time sufficient SEB is present to stimulate adequate cytokine production.

The experiment with antibodies to interleukin-2, which was not reported in the original paper, is difficult to interpret without positive controls. Were the antibodies effective? If lymphokine-withdrawal apoptosis did take place, would it have been detected? The failure to detect active SEB 36 hours after injection also lacks a positive control; could Ochi and Kawabe, for example, detect SEB 6 hours after injection? In any event, the concentration of SEB that is important is not what can be detected in an arbitrary assay but what is effective *in vivo*. We still believe¹ that the loss of $V\beta 8^+$ cells *could* be by activation-induced death, but that the evidence is not convincing.

J. JOHN COHEN
RICHARD C. DUKE
KAREN S. SELLINS

Department of Microbiology and
Immunology,
University of Colorado Health Sciences
Center,
Denver, Colorado 80262, USA

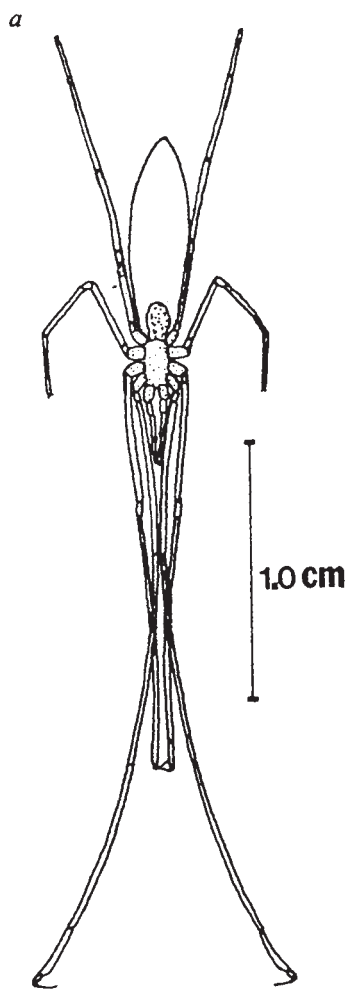
1. Cohen, J. J., Duke, R. C. & Sellins, K. S. *Nature* **352**, 199–200 (1991).

2. Kawabe, Y. & Ochi, A. *Nature* **349**, 245–248 (1991).

3. Duke, R. C. & Cohen, J. J. *Lymphokine Res.* **5**, 289–300 (1989).

Impaled prey

SIR — Remarkable evolutionary innovations are characteristic of species on isolated land masses. This is particularly true of the Hawaiian Islands, where the processes of evolution and extinction have been accentuated and accelerated, yet many species remain unknown and undescribed¹. The effects of anthropogenic disturbance are similarly acute. As a consequence, the archipelago represents a microcosm for global issues and concerns in evolution and conservation.



a, *D. raptor* at rest, ventral view, showing the long prolateral claws of leg pairs I and II, and the typical resting posture. b, Tarsus and claws of right leg I, showing elongation of the prolateral claw (length 0.4–0.7 mm; 45–60% length of tarsus). The relatively small retrolateral and medial claws, which barely project beyond the limits of the articulated base, have the standard proportions for these claw for Tetragnathidae. Strong macrotrichiae line the ventral margin of the distal 2/3 of the tarsus. $\times 225$.

Striking examples of adaptive radiation in the Hawaiian Islands have been documented in the genus *Drosophila*, the land snails and the honeycreepers². The most recent species radiation to be discovered in the archipelago is a lineage of diverse, conspicuous and abundant spiders in the family Tetragnathidae³. In particular, one of the most remarkable morphological features ever found in spiders (immense elongation of the tarsal claws) is exhibited by the endemic Hawaiian tetragnathid *Doryonychus raptor* Simon: the prolateral claws on the tarsi of leg pairs I and II of *D. raptor* are immensely elongated (see figure), and their spinneret morphology indicates

that the species has lost the ability to spin capture webs. *D. raptor* captures insects using only the enlarged prolateral claws to impale prey and draw them to the chelicerae in a single, rapid movement. The spider is strictly nocturnal, spending most of the activity-period hanging upside down from silk threads. Small insects are snagged directly from the air using a single long claw. For larger insects the spider uses both long claws on legs I, or sometimes all the long claws. On securing the insect in the chelicerae, the claws are immediately

withdrawn, but used again at intervals during feeding, to move, and finally discard, the insect.

The highly specialized mode of foraging in *D. raptor* may underlie the extreme restriction of its distribution⁴. The species is almost entirely confined to small, remnant pockets of lowland forests directly below high waterfalls on the Hawaiian island of Kauai. The unique specialization and range restriction exhibited by *D. raptor* render it a valuable candidate for studies of conservation on the basis of charismatic value alone. In addition, being a generalized predator of herbivores, it possesses considerable ecological value⁵. But the true potential of the system lies in the extremes it represents, allowing it to serve as a model for global concerns in conservation biology. Processes of evolution² and anthropogenically induced extinction¹ have occurred on a heightened scale in the Hawaiian archipelago; nowhere are these processes better exemplified than in the case of *D. raptor* in the vanishing lowland forests of Kauai. Consequently, quantitative assessment of such effects as

alien invasion and gene-pool reduction will be particularly instructive. It is important that model systems such as those in Hawaii be studied before they disappear.

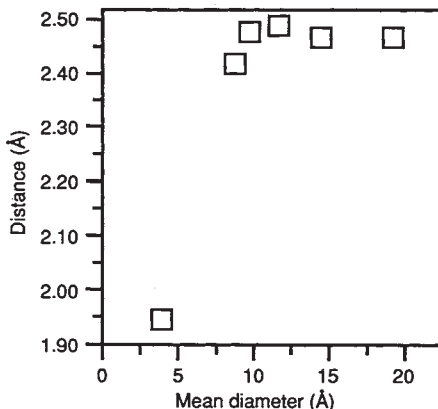
ROSEMARY G. GILLESPIE

*Hawaiian Evolutionary
Biology Program,
University of Hawaii,
Honolulu, Hawaii 96822, USA*

1. Howarth, F. G. & Ramsay, G. W. *The Conservation of Insects and Their Habitats* (eds Collins, N. M. & Thomas, J. A.) 71–107 (Academic, London, 1991).
2. Simon, C. *Trends Evol. Bio.* **2**, 175–178 (1987).
3. Gillespie, R. G. *J. Arachnol.* (in the press).
4. Pianka, E. R. *Evolutionary Ecology* 3rd edn (Harper and Row, New York, 1983).
5. Pimm, S. L. & Gilpin, M. E. *Perspectives in Ecological Theory* (eds Roughgarden, J., May, R. M. & Levin, S. A.) 287–305 (Princeton University Press, 1989).

Crystal clusters

SIR — In my recent News and Views article¹, I described the unusual behaviour of extremely small molybdenum crystallites: groups of tiny cubes are capable of self-assembling into rather larger cubes. I raised the question of whether the larger cubes thus assembled



Near-neighbour distance versus cluster size.

have the physical properties characteristic of the smaller size of crystallites or those characteristic of the assembled cubes. One such property, I pointed out, would be the lattice parameter, which might be expected to vary with crystallite size, but at that time I knew of no measurements of lattice parameter as function of size for very small crystalline metal clusters.

P. Haubold of the University of Saarbrücken has told me that related experimental evidence is available. Several physicists have determined interatomic spacings in metallic clusters by means of EXAFS (extended X-ray absorption fine structure) measurements, and have indeed been able to establish that this spacing decreases with cluster size^{2,3}.

The pattern of packing may alter (for example, iron clusters are body-centred cubic down to about 9 Å, but for smaller sizes the packing becomes closer). The oscillations of absorption coefficient near an X-ray absorption edge (EXAFS),

after Fourier inversion, yield interatomic separations irrespective of crystal structure. The pair-distribution function changes radically for very small, imperfectly crystalline clusters. The figure, from ref. 2, shows how nearest-neighbour distances, deduced from EXAFS, change as a function of cluster size for iron. If molybdenum, another body-centred cubic metal, behaves similarly, then one would expect very little change in interatomic spacing (hence lattice parameter) between the small cubes of edge length about 50 Å and the assembled cubes of an average of 175-Å edge length; even the small cubes are not quite small enough for the interatomic spacing to change radically.

One wonders whether clusters so small that they do not have a proper crystal structure (≤ 10 Å) can self-assemble to form larger clusters which are crystalline.

ROBERT W. CAHN

*Department of Materials Science and
Metallurgy,
University of Cambridge,
Pembroke Street,
Cambridge CB2 3QZ, UK*

1. Cahn, R. W. *Nature* **353**, 208 (1991).
2. Zhao, J. et al. *Proc. Mat. Res. Soc. Symp.* **143**, 151–156 (1989).
3. Montano, P. A. et al. *Phys. Rev. Lett.* **56**, 2076–2079 (1986).

No planet orbiting PSR1829–10

SIR — In an earlier paper¹, we reported a cyclic variation in the arrival times of the pulses from the neutron star PSR1829–10 with a period close to 6 months, and presented this as evidence for a 10-Earth-mass planet. As we noted in that paper, we were concerned that the 6-month periodicity might be an artefact concerned with the Earth's orbit around the Sun, but were encouraged by the fact that no such periodicity appeared in observational data for the 300 other pulsars currently under observation. We have nevertheless re-examined the algorithm used in compensating for the Earth's orbital motion and now find that we can account for the observed radiation without the presence of a planet.

The standard analysis (p. 105 of ref. 2) involves correcting the observed arrival times to the barycentre of the Solar System using a precise ephemeris for the position of the Earth. An analytical model for the pulsar rotation and position is then adjusted to minimize a set of residuals, the differences between the observed barycentric arrival times and model times. Because this is a differential process, the approximation is made that the orbit of the Earth is circular. Provided that the difference between the

position used for the barycentric correction and the model position is small, this is perfectly valid. However, in the case of PSR1829–10, we have found that the error in the initial position used to compute the barycentric arrival times was unusually large, amounting to 7 arcmin from the final fitted position; the ellipticity of the Earth's orbit then becomes important, causing the appearance of a 6-month sinusoidal oscillation of 8-ms amplitude in the residuals. It is usual to keep the position used for the barycentric correction close to the fitted position, but unfortunately that was not done in this case. We have reprocessed the data using the improved position of the pulsar in calculating the barycentric correction, and find that the pulsar has no companion body.

Our failure to recognize the result of a position difference, and to perform the usual procedure, has resulted in the apparent planet, and we must accept full responsibility for this error. Although there is no planet orbiting PSR1829–10, pulsar timing remains the most sensitive technique for detecting planets outside the Solar System, as exhibited in last week's paper by Wolszczan and Frail³.

A. G. LYNE

M BAILES

*Nuffield Radio Astronomy Laboratories,
University of Manchester,
Jodrell Bank, Macclesfield,
Cheshire SK11 9DL, UK*

1. Bailes, M., Lyne, A. G. & Shemar, S. L. *Nature* **352**, 311–313 (1991).
2. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
3. Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).

Intraplate not always stable

SIR — Matters of semantics often seem trivial, but are sometimes necessary to avoid conveying misleading information. In a recent paper¹ describing the faulting associated with the Ungava (Quebec) earthquake of 1989, the authors assert (and Crone² repeats) that “[w]orldwide, only ten historical intraplate earthquakes are known to have produced surface faulting,” citing Johnston and Bullard³. What we actually said was that “including Ungava there are only 10 documented examples of surface rupture from historic earthquakes in the stable continental regions of the earth.”

Why bother with a distinction between intraplate and stable continental regions (SCR), as both terms refer to the interiors of tectonic plates, as opposed to their boundaries? One reason — with more fundamental reasons to follow — is that substituting intraplate for SCR renders the quoted ten surface faulting events at least an order of magnitude too

low. 'Intraplate' would incorporate continental region of active tectonics such as south central Asia, in which surface rupture is relatively abundant. In the western United States alone, one can cite ten examples of intraplate surface-faulting earthquakes outside the San Andreas fault system.

Continental crust covers an area of $2.1 \times 10^8 \text{ km}^2$, of which 'stable' crust (the Precambrian cratons and shields, Palaeozoic mountain belts and margins no younger than Palaeogene ($\sim 37 \text{ Myr}$)) comprises $\sim 1.32 \times 10^8 \text{ km}^2$ (63%). Tectonically active crust, in which large-scale faulting, mountain building or magmatic activity extends into the Cenozoic era (0–65 Myr) covers $\sim 7.3 \times 10^7 \text{ km}^2$ (35%), with submerged continental plateaux providing the remaining 2%.

This division of continental crust into active and stable components⁴ is certainly not the only one possible, and the term 'stable' is only relative, as large earthquakes do occur in SCR crust. But the need for some such distinction is highlighted by the great range of surface faulting rates observed in continental regions. An alternative to using the tectonic age of the crust as a means of distinguishing relatively active from relatively stable regions is just becoming available from the satellite geodetic techniques of VLBI⁵ (very-long-baseline interferometry), SLR⁶ (satellite laser ranging) and GPS⁷ (Global Positioning System). These and other preliminary results suggest that SCR deform at a strain rate of 10^{-9} – 10^{-10} yr^{-1} — orders of magnitude less than those that characterize active intraplate (10^{-7} – 10^{-9} yr^{-1}) and plate-boundary zones (10^{-5} – 10^{-6} yr^{-1}). As one would expect, these large differences in strain rate are reflected in the observed seismicity⁸.

Stable continental (and oceanic) crust is our closest approximation to the plate-tectonic ideal of monolithic plates interacting, reacting and deforming only at their boundaries. This is not true of all intraplate regions. In cases, such as the study of surface faulting, where the term 'intraplate' combines two quite different species of crust, it should be avoided.

ARCH C. JOHNSTON

Center for Earthquake Research,
Memphis State University,
Memphis, Tennessee 38152, USA

1. Adams, J., Wetmiller, R. J., Hasegawa, H. S. & Drysdale, J. *Nature* **352**, 617–619 (1991).
2. Crone, A. J. *Nature* **352**, 570 (1991).
3. Johnston, A. C. & Bullard, T. *Seismol. Res. Lett.* **61**, 152–153 (1990).
4. Johnston, A. C. & Kanter, L. R. *Scient. Am.* **262**(3), 68–75 (1990).
5. Ward, S. N. *J. geophys. Res.* **95**, 21965–21981 (1990).
6. Smith, D. E. et al. *J. geophys. Res.* **95**, 22013–22041 (1990).
7. Billiris, H. et al. *Nature* **250**, 124–129 (1991).
8. Johnston, A. C. in *Earthquakes at North-Atlantic Passive Margins: Neotectonics and Postglacial Rebound* (eds Gregersen, S. & Basham, P. W.) 299–327 (Kluwer, Dordrecht, 1989).

Methane and Milankovitch cycles

SIR — Crowley suggests¹ that the northern summer radiation maxima may have partly controlled atmospheric methane levels² by enhancing methane output by high-latitude ecosystems. But this idea neglects what is known of major controlling factors on high-latitude production.

At present, the high-latitude wetlands rely for their capacity for methane production on the anaerobic, waterlogged conditions which are associated with, and strongly reinforced by, extensive peat deposition³ (and Clymo, personal communication). From palaeoevidence⁴ it seems that an extensive high-latitude complement of anaerobic peat wetlands would have taken millennia to form, after the climate initially became suitable for them. For example, when considering the late glacial-early Holocene methane rise and peaks (12 and 9 kyr), the necessary anaerobic conditions may not have been extensive in the high northern latitudes at the time of the summer maximum.

The main Siberian peat-basins and other wetlands have progressively spread from very localized nuclei to form an extensive blanket during the Holocene⁴. The greater growing season evaporation potential and a resulting summer lowering of water tables may well have favoured less waterlogging at the summer radiation maximum. Large areas of the high latitudes of the Northern Hemisphere that are currently covered by peatlands and are now major methane producers were still under ice at the times of the Early Holocene methane rise and peaks (particularly

the 12-kyr peak). In areas that the ice had already vacated, the generalized picture of deglaciating landscapes being pock-marked by lakes and dissected by meltwater streams does not necessarily point to extensive anaerobic conditions.

In northwestern Europe, sedimentary evidence suggests that the late-glacial and early Holocene (12–9 kyr) landscapes were lacking in extensive anaerobic and peat-forming conditions. Only as the Holocene proceeded did extensive bogs begin to form naturally in this area. Thus, we should perhaps think in terms of the main development of the northern high-latitude methane source having a lag period, which means that it is unlikely to reach its maximum potential for thousands of years, until after each insolation peak has passed.

J. M. ADAMS

Department of Plant Sciences,
University of Cambridge,
Downing Street,
Cambridge CB2 3EA, UK

H. FAURE
N. PETIT-MAIRE

Laboratoire de Géologie
du Quaternaire,
CNRS, Case 907,
Luminy, F-13288 Marseille
Cedex 9,
France

1. Crowley, T. J. *Nature* **353**, 122–123 (1991).
2. Chappellaz, J., Barnola, J. M., Raynaud, N., Korotkevich, Y. S. & Lorius, C. *Nature* **345**, 127–131 (1990).
3. Matthews, E. & Fung, I. *Global Biogeochem. Cycles* **1**, 61–86 (1987).
4. Neustadt, M. I. in *Late Quaternary Environments of the Soviet Union* (ed. Velichko, A. A.) 201–206 (Longman, London, 1984).

Patterns of population spread

SIR — The recent News and Views article¹ accompanying our paper², and various press reports about our work, could have provided a false impression about earlier work. As we pointed out², the hypothesis that agriculture in Europe spread by demic diffusion was first proposed by Ammerman and Cavalli-Sforza^{3–5}. The theory predicts gene-frequency clines in an approximate northwest-southeast direction in Europe. Cavalli-Sforza's group first described⁶ such clines based on synthetic variables obtained by principal components analysis, but they did not test the statistical significance of these trends. Our laboratory first tested the statistical significance of patterns by means of spatial autocorrelation analysis. We confirmed the existence of northwest-southeast clines for gene frequencies but not for cranial variables. Our latest paper² reported a method for testing the genetic evidence against the spe-

cific pattern of neolithic spread.

ROBERT R. SOKAL
CHESTER WILSON

Department of Ecology and Evolution,
State University of New York,
Stony Brook, New York 11794-5245,
USA

NEAL L. ODEN

Department of Preventive Medicine,
Health Sciences Center,
State University of New York,
Stony Brook, New York 11794-8036,
USA

1. Jones, J. S. *Nature* **351**, 97–98 (1991).
2. Sokal, R. R., Oden, N. L. & Wilson, C. *Nature* **351**, 143–145 (1991).
3. Ammerman, A. J. & Cavalli-Sforza, L. L. in *The Explanation of Culture Change* (ed. Renfrew, C.) 343–358 (Duckworth, London, 1973).
4. Ammerman, A. J. & Cavalli-Sforza, L. L. in *Transformations: Mathematical Approaches to Culture Change* (eds Renfrew, C. & Cooke, K. L.) 275–294 (Academic, New York, 1979).
5. Ammerman, A. J. & Cavalli-Sforza, L. L. *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton University Press, 1984).
6. Menozzi, P., Piazza, A. & Cavalli-Sforza, L. L. *Science* **201**, 786–792 (1978).

The paradox of genius

Stephen Jay Gould

Darwin. By Adrian Desmond and James Moore. *Michael Joseph*: 1991. Pp. 808. £20. To be published in the United States by Warner Books later this year.

THE botanist J. D. Hooker, Charles Darwin's closest friend and confidant, contributed a short preface to the great 1909 Cambridge Festschrift commemorating both the fiftieth anniversary of the *Origin of Species* and the hundredth of Darwin's birth. Hooker, at the age of 92, recalled his pleasure in Darwin's trust and cited the volume of correspondence that he had received from Darwin alone: "upwards of a thousand pages of foolscap, each page containing, on an average, three hundred words."

Darwin lived on that precious nineteenth-century crux of maximal information for historians — close enough to the present to permit preservation and not too close for the passage of most communication through electronic channels into permanent oblivion. Beyond the range of a shout, every scrap of information then passed in palpable written form.

With Darwin we are doubly blest. He not only lived at the right time but was also a compulsive keeper of records and tabulations. We even know the scoresheet on his nightly backgammon with his wife Emma after nearly 6,000 contests at two per evening (Charles was up by 305 games in January 1876).

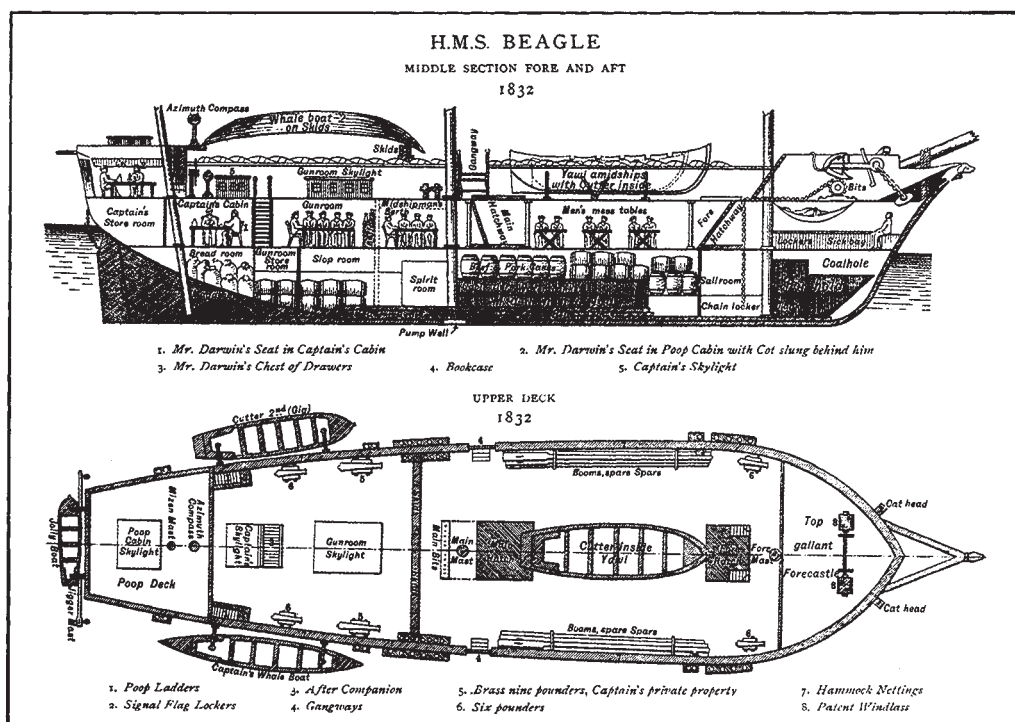
The combination of Darwin's seminal importance in the history of human thought with maximal evidence about his life has spawned a historiographical effort that professionals call the "Darwin industry", both in admiration for its prodigious accomplishment and slight amusement at its obsessive probing. Adrian Desmond and James Moore are industry standards, and their ample biography, unquestionably the finest ever written about Darwin, is a bright (if literally ponderous) testimony to the triumph of this particular industrial revolution.

I was beginning my own higher education when the industry received its jumpstart at the worldwide centennial celebrations for the *Origin of Species* in 1959 — a start inspired, in large part, by the 'discovery' of the extensive notebooks and private jottings that Darwin

kept throughout his life, especially as a young man setting out to reform the entire intellectual world. (These documents had always been there, but were ignored and therefore unknown and invisible.) I have lived ever since in the penumbra of this industry as a practising

recognition that, although nature surely embodies factual truth accessible to human investigation (radical historians might even demur at this evident and necessary proposition), science does not and cannot 'progress' by rising above social embeddedness on wings of a timeless and international 'scientific method'. Science, like all human intellectual activity, must proceed in a complex social, political and psychological context: greatness must therefore be grasped as fruitful use, not as transcendence.

Darwin, so often depicted as a positivist hero, self-exiled at Downe in Kent amidst his plants and pigeons, emerges



Darwin's home afloat — plans of the ship that gave him passage to a new world. (From *Darwin*.)

scientist, darwinian and sometime historian. When I think back to what most people knew of Darwin and thought about the history of science just before 1959, I am astounded at the salutary changes in our understanding, so well embodied in this fine book.

Preindustrial history of science was strongly positivist, with emphasis on individuals whose superior skills in observation and reasoning could break through social prejudice to reveal nature's objective truths. In this context, biographies of Darwin tended to be either hagiographical (when written by scientists who applauded such independence from social shackles) or critical (when written by humanists who decried Darwin's supposed literary and cultural benightedness, and who therefore blamed him directly for all subsequent social misuse of his ideas).

The history of science has been transformed during the past 30 years by a

from his industrial reconstruction as the finest example of a man who could succeed only by virtue of his social position and of the political changes, so fortuitous from the standpoint of his power to alter them, that convulsed England during his career. Darwin has now become, and properly, the quintessentially socially embedded scientist. Desmond and Moore are brilliant in their relentless and integrative pursuit of this truly unifying theme. (I must record as my only general criticism of the book some unhappiness at the drastic curtailment of nearly all traditional material on the actual content of Darwin's scientific ideas. We never even get a coherent description of natural selection. For some of the minor books, we learn about print-runs and Darwin's tummy troubles as he faced the proofs, but virtually nothing about their content. As Darwin knew so well, good revolutions transmute, but do not entirely raze, older

ways. Still, I can hardly suggest that *Darwin* be lengthened.)

Charles Darwin was a man in the middle, politically and socially, and this position set and conditioned all his strategies. His early teachers and benefactors were Anglican Tories, the gentlemen and ecclesiastical naturalists of Oxbridge. (He would have become a parson naturalist, had not another aristocratic Tory, Captain FitzRoy, intervened and offered him passage to a new world on the *Beagle*.) These men maintained an unshakeable conviction that our ordered and harmonious Universe required the direct and continuous superintendence of a providential God. Otherwise, matter would devolve to chaos and formlessness, just as society would crumble without direct, intelligent government by the upper classes. Anti-evolutionism was the core of a worldview, not a question for empirical resolution.

Darwin was, by both family tradition and personal conviction, a Whig — defender of a broad range of freedoms, from the abolition of slavery to the least-trammelled of *laissez-faire* markets with all the malthusian consequences for those who fail. (We may safely brand as utter nonsense the common claim that Darwin happened to pick up Malthus's *Essay on the Principle of Population* for fun one day and thus came by chance upon the key insight of natural selection. Malthus stood at the centre of his intellectual world; the most strident of English malthusians, Harriet Martineau (who won great respect and attention in this gynophobic society), was a constant companion of Erasmus Darwin during the early post-*Beagle* years, when Charles hung out at his brother's *soirées*.)

But genuine radicals haunted the other extreme, and their largely ephemeral press (so brilliantly reconstructed and dissected by Desmond in his earlier book, *The Politics of Evolution*) invoked evolution as an underpinning of reform. Not Darwin's whiggish natural selection to be sure, but a lamarckian doctrine of universal progress, inherent in matter and propelled by self-effort. Society, if unfettered, works *sui generis* and does not need Paley's watchful God to stir it into motion or keep it in order.

Darwin had developed his theory of natural selection by the end of 1838. What was he to do? He was an upper-class Englishman through and through, on the verge of marriage, with a passion for natural history and no desire to work at an ordinary job. (He became a country squire of much greater wealth than I had ever realized before reading this book — a fortune grounded in inheritance, but substantially augmented by

shrewd investment in land and railways.) He was a philosophical radical in many ways, and he lost all vestiges of conventional religious faith after the death of his favourite young daughter in 1851. But he was socially conservative and conventional almost to an extreme; he hated the radicals as much with his heart as his mind. He could not bear to abet a social transformation that he despised. He had an idyllic marriage with a wonderful and brilliant woman (too invisible in this book and so much deserving her own biography), who had studied piano with Chopin, learned to speak several languages, and then devoted her life to succouring Charles and raising their family. Emma was devout and wept for her husband's religious doubts (while supporting all his work unstintingly, down to the details of reading proofs).

How could Darwin proceed with his beloved brainchild — a tenable theory of evolution that (as he knew so well, for he was not a modest man) would change the world irrevocably? He retreated to illness and other projects, including an eight-year study of the taxonomy of barnacles. How can we begin to understand this cardinal story of his intellectual life — his 20-year delay from devising the theory in 1838 to his flushing out by Wallace in 1858 — without this intricate social, political and personal context?

Partial answers

The answer is that we simply cannot, and that all previous standard explanations are either nonsense or hopelessly partial. He was not delaying for the good-old positivist reason of needing time to collect his facts. (He gathered acres of facts, and they did win his case, but his primary motive was fear.) He was not afraid to expose evolution, for no heresy was more common among nineteenth-century biologists. He did fear the social consequences of the materialist outlook inherent in his own version of malthusian change, a vision so ripe for radical exploitation. He could not dream of exposing such dynamite in the 1830s or early 1840s when a politically troubled England, convulsed by Chartism and agitation over the poor laws, verged on potential revolution.

But a different and seemingly permanent sun shone upon the late 1850s. Both England and Darwin were now comfortable, wealthy and secure. Crystal Palace marked the triumph of victorian industry and growing imperial domination. A new force of ambitious younger scientists had emerged at centres of power in London, and Darwin became their hero and (usually silent) elder statesman. Hooker, Huxley, Tyndall and Darwin's neighbour Lubbock (a scientist and member of parliament with valuable

connections) all stood for an accessible, modernized, paid, professionalized science against the old, closed world of amateur aristocrats and parsons. Natural selection was the right doctrine for a new, secular, urbanized, industrial nation.

I knew the outlines of this social context before reading *Darwin*. But history is narrative and true understanding lies in integrating the details. I thank Desmond and Moore most of all because I think that I have now finally grasped the outlines of why Darwin ticked in his century. (The details of history can be so daunting, but there are no short cuts. No simplifying laws of nature will render the main puzzles of Darwin's life. Like all people, he was a historical item living in a complex time, and proper explanations lie in the fabric of detail. The saving grace rests with the beauty of the fabric. But I was astounded to note over and over again that single paragraphs and sentences of this ample work have been the basis for entire tomes by others. The death of Darwin's mother, rendered here in half a paragraph, is the foundation for Bowlby's recent and most interesting treatment of Darwin's lifelong illness. Less than a page on the "principle of divergence" caused me months of study, resulting in a 100-page chapter in a forthcoming book.)

But what made Darwin tick? Why him? For all Desmond and Moore have helped me to understand, their social approach leaves me with an immensely heightened sense of paradox. For I now grasp why Darwin was absolutely the wrong person for the job. He was exactly what Huxley yearned to sweep out — an amateur upper-class naturalist of inherited wealth, an Oxbridge product who had collected beetles and hoped for a parsonage. He was groomed by the Anglican naturalist-divines Henslow and Sedgwick, and he never broke away from their social style.

I can locate pieces of the answer. Darwin had an uncanny appreciation of the elusive but necessary middle-ground of all intellectual revolution — the sharp precipice between valleys of crankiness and conventionality. He had an obsessive inquisitiveness (and honesty in viewing results thereby obtained) that can only be called fierce. His questionnaires and experiments were models of badgering thoroughness: he could not bear any slippage of intellectual loyalty from supporters. I cannot integrate all the pieces that, in the final paradoxical analysis, make Darwin the disproof of any sociocultural determinism — but I think we usually call the totality 'genius'. □

Stephen Jay Gould is at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

The turn of the screw

Jonathan Slack

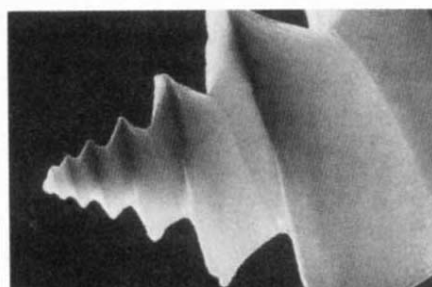
Biological Asymmetry and Handedness, Ciba Foundation Symposium 162. Edited by Gregory R. Bock and Joan Marsh. Chairman: L. Wolpert. Wiley: 1991. Pp. 327. £39.50, \$69.50.

LEWIS Wolpert is left-handed, but we know that he is not made of antimatter, neither are his proteins composed of D-amino acids. He does not even have *situs inversus viscerum*, the rare condition in which the internal organs are arranged as a mirror-image of the normal. So at what level is his own bilateral symmetry broken, and how does it happen? Wolpert has long been concerned with the general issue of why there are left-right asymmetries in organisms such as vertebrates, which seem to a good approximation to be bilaterally symmetrical. This book, the proceedings of a Ciba Foundation Symposium held last year, investigates the problem in detail and considers every level of asymmetry from nonconservation of parity in the weak nuclear interaction to laterality of human motor functions.

The story starts with life on Earth and why all proteins are composed of L-amino acids. According to the principle of CP conservation, the true mirror-image of a chiral molecule must be made of antimatter. The terrestrial enantiomers, the D-amino acids, are slightly less stable in the protein alpha-helix or beta-sheet than are the L-forms. Although this difference is not measurable, it is sufficient to lead to the extinction of the D-forms in the primaeval soup of the early Earth if we assume that there was an autocatalytic stereospecific synthesis of the amino acids from achiral precursors. Of course we cannot exclude the possible additional effects of asymmetrical environmental influences such as circularly polarized sunlight.

Having got L-amino acids (and D-sugars, which are also CP-favoured), do all higher asymmetries flow inexorably from them? Probably not. It does seem that the main themes of protein secondary structure would be reversed if the handedness of the monomers were reversed. This might carry us up as far as large helical structures such as viruses, but the arrow of causality becomes lost in the more complex three-dimensional patterns that are required for packing and recognition of globular proteins. At this level, the same structure can be made in many different ways. For instance, if we make an antibody to a biologically active molecule and then

make an antibody to the first antibody, the second antibody (an 'anti-idiotypic') sometimes shares the biological activity of the original antigen. But the primary sequence of the anti-idiotypic will be different from that of the original antigen. So it is true that three-dimensional structure is determined by primary structure, and asymmetries in this structure will be determined by the asymmetry of the monomers, but there are so many ways of making any required structure that higher asymmetries could be created from any kind of basic monomer. At the meeting, this became known as the



Mollusc shells of the right-handed persuasion. From the third edition of *The New Ambidextrous Universe: Symmetry and Asymmetry from Mirror Reflections to Superstrings* by Martin Gardner. Published by Freeman, £10.95 (pbk), £15.95 (hbk).

'Choithia gap', but it is also the familiar epistemological gap that exists in embryology between understanding structures that arise from molecular self-assembly and understanding those that require higher levels of causation.

The next theme of the book concerns the symmetry of embryos. Some embryos develop without bilateral symmetry from the outset. It is well known that the left-handed adults of the snail *Lymnaea* begin their lives with a laevotropic rather than the normal dextrotropic spiral cleavage. But it is less well known that the veligar larvae of gastropods are at least roughly bilaterally symmetrical, and the connection between the sense of early cleavages and the sense of the shell gland torsion is not understood. Some embryos acquire an early bilateral symmetry, probably because this is an automatic consequence of having two signalling centres in a cell mass such as the vegetal pole and the dorsal lip of an amphibian blastula. There is much evolutionary speculation about which axes

or planes of symmetry in vertebrates are homologous to those in protosome invertebrates, and whether bilateral symmetry or asymmetry is primitive, a discussion that seems to me to be blissfully unencumbered by data.

In vertebrate embryos, there are ways of reversing the left-right asymmetry of the viscera. Fifty per cent of *Xenopus* embryos have *situs inversus* if they are treated during gastrulation with agents that interfere with cell contacts and adhesion, such as RGD peptides, heparinase or inhibitors of glycosylation. The same proportion of mice homozygous for the mutant *iv* gene have *situs inversus*. Also, about half of those humans homozygous for a related disorder, Kartagener's syndrome, associated with a defect in the dynein arms of cilia, may have *situs inversus*. The molecular causes of these changes are probably diverse, and may not necessarily be closely related to the normal breaking of bilateral symmetry in development. But these examples do show that inhibition of the normal mechanism does not reverse the normal asymmetry; instead it causes a loss of preference for the sense of asymmetry. Various models for normal development are discussed. All seem to depend on an asymmetrical molecule that can be aligned with the body axes and hence allow different processes to occur on right and left sides. Whether this really represents a bridge from molecular to organismic asymmetry depends on how big the asymmetrical molecule is, or, to put it another way, on which side of the Choithia gap it lies.

Yet another epistemological gap may separate the asymmetries of vertebrate embryos from asymmetries of cortical function in postnatal life. Anatomical asymmetries in the human brain are apparently more pronounced than those of other mammals. The prevalence of left-handedness is about 8 per cent in the overall human population, rising to 26 per cent in offspring of left-left matings — as usual in human biology there seems to be some genetic input, but it is not clear-cut. Despite inherited or acquired preferences for the use of a particular organ, these can later be overridden, as in people who have lost the use of their hands but have learned to draw using their mouth or toes.

I still suspect that the small perturbations likely to be responsible for the breaking of left-right symmetry are not very specific or easy to investigate. But fortunately, no book that jumps from spiral staircases to flatfish to dynein arms in a few pages can fail to be stimulating.

Jonathan Slack is in the ICRF Developmental Biology Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Science for the citizen

Andrew Clifford

Strange Weather: Culture, Science, and Technology in the Age of Limits. By Andrew Ross. Verso: 1991. Pp. 275. \$59.95, £32.95 (hbk); \$16.95, £10.95 (pbk). Distributed in North America by Routledge, Chapman and Hall.

SCIENCE informs almost all aspects of everyday life. Its method and metaphors are part of ordinary men and women's conversations about everything from cooking to football to sex. Between professional scientists and the public are discreet and intricate minicultures, such as those of 'new agers', 'cyberpunks' and computer hackers, which locate themselves around the margins of scientific practices and theories; and just as establishment science (as Ross calls it) participates in the ambitions of the political and economic mainstream, so these alternative and often adversarial groupings are to some extent the vanguard of the scientific nonestablishment — that is, the public. *Strange Weather* is mainly a study of these minicultures, in particular of how they open up and close off science to ordinary people.

Ross's left-libertarian journey begins well with careful studies of individual groups, but too often weighs itself down with apocalyptic and overweening political 'recommendations'. Yet, despite his jaundiced view of the scientific method as being no more valid and every bit as political as any other way of interacting with the world, his perspective does allow him some original insights.

He begins by scrutinizing the new-age movement, moving on to examine computer hacking, cyberpunk, futurology and global warming. But it is earlier on that he scores most points. In looking at the new-age movement, for example, he points out that although it "dreams us back to the prescientific and to the alchemist's kitchen", it has also assumed "a virtuoso, experimental role in reconstructing a humanistic *personality* for science": it is, in other words, a rather muddled attempt to restore science in some way to the people. Similarly, Ross examines computer hacking as a form of rebellion. Although unimpressed by 'celebrity' computer hackers, he argues that there are other forms of hacking (which he terms "reconstructive" as opposed to "deconstructive") that convey a lot about their perpetrators' feelings of impotence towards technology (such as the petty sabotage of operating systems by office workers).

Of course, the greater the democra-

tization of popular science, the more Ross is pleased, sabotage and theft or not. The strengths and weaknesses of this position are epitomized in the final, largest section on global warming and the new "global weather culture". Ross provides a telling analysis of the Weather Channel on US television, where countries sharing the North American continent appear on the map only "when 'their' weather is seen to be affecting 'ours' . . . On the weather map, Cuba, for the most part, still does not 'exist'; it does not influence US weather, since the US has no influence over it." He goes on to note that "Instead of feeling the weather as we have felt it historically, as part of a shared local or even national culture, we are encouraged to think of it globally. On the one hand, this will promote our attention to the interdependence of environmental factors across cultures and continents; on the other, it may help to dissipate people's faith in the efficacy of local action."

But if this is true, it's also not true enough. Neither politics nor ecology is

an either/or process. It is oversimplifying the situation to argue, as Ross does, that the international call for 'humanity' to be more ecologically responsible conceals and denies the more specific responsibility of large institutions and corporations (although he does not deny the need for large-scale solutions). And far too often his observations metamorphose into guidelines for future political strategies. Worst of all, he really does seem to believe that if power were in the hands of "the people", everything would become ecologically hotly hot.

Indeed, so generalized and proletarian does his argument become that he gradually succumbs to the same globalist and omnipotent tendencies as the corporations he believes have done so much harm. Even so, the book contains much food for thought on how science influences everyday life and thus on how all of us cook, play football and have sex. □

Andrew Clifford, a freelance writer on cultural trends, is at 25 Keswick House, Crawford Road, London SE5 9NL, UK.

Geological miracles

Paul G. Davis

Santana Fossils: An Illustrated Atlas. Edited by J. G. Maisey. T.F.H. Publications: 1991. Pp.459. £125, \$200.

DEPOSITS characterized by exceptionally preserved fossil faunas, otherwise known as Lagerstätten, are of immense geological importance. The fossil record generally preserves only hard parts of animals and rarely preserves soft tissues, which normally decay without trace. So Lagerstätten give a complete and more accurate picture of diversity within the fossil record (and show just how many animals are not preserved under normal circumstances) as well as yield information about palaeobiology.

The Santana formation in Brazil consists of a variety of strata (from evaporites to lithographic limestones to shales with nodular concretions) housing a host of wonderfully preserved fossils. These Lagerstätten represent a unique 'snapshot' of life during the early Cretaceous period of 145–65 million years ago. Some animals were instantaneously fossilized within the formation (the so-called 'Medusa effect'). Indeed, many of the fish have been so rapidly and perfectly preserved in calcium phosphate that structures such as nuclei of muscle fibres, secondary gill lamellae and stomach wall fragments can still be seen.

The chapters cover basic continental drift, basic cladistics, the sedimentology,

stratigraphy and palaeogeography of the Araripe basin, the placement and relevance of this basin in the opening of the South Atlantic, taphonomy of the fossils (strangely hidden under the title of "Fossil Forensics"), their local collection and preparation, and advanced preparation techniques for the laboratory. The most extensive chapter, which fills most of the book, is a systematic atlas of the fossils.

The formation is renowned for its fossil fish, the most spectacular of which are the three-dimensional specimens that are found in nodules of calcium carbonate. Owing to their aesthetic appeal and large-scale collection by local villagers, these fossil fish can now be found in shops around the world. The descriptions of these fossils fill more than 200 pages of diagnosis, diagrams and colour photographs (it is a pity that some lack scale bars), and are obviously intended for the more serious worker or anyone trying to identify new specimens. The rest of the atlas chapter covers other fossil vertebrates such as pterosaurs (providing a useful English summary of all the German literature on these flying reptiles), frogs, turtles, dinosaurs and crocodiles; the invertebrate fauna, which includes echinoderms, insects, spiders, crustaceans and molluscs; and the fossil plants.

The book brilliantly bridges the gap between semipopular and academic works. In particular, the many high-quality colour photographs make it an especially pleasurable read. □

Paul G. Davis is in the Department of Geology, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK.

Chromatin as an essential part of the transcriptional mechanism

Gary Felsenfeld

The increasingly detailed biochemical definition of the protein complexes that regulate gene transcription has led to the re-emergence of questions about the role of histones. Much recent evidence suggests that transcriptional activation requires that transcription factors successfully compete with histones for binding to promoters, and that there may be more than one mechanism by which this is achieved.

IN the days before much was known about gene expression in eukaryotes, histones were the most obvious candidates as regulators of transcription, though it was not clear how they could be anything but nonspecific repressors. With the discovery of *cis*-regulatory elements and *trans*-activating factors, the histones were temporarily forgotten, despite the realization that the packaging of DNA in nucleosomes and further into 30-nm chromatin fibres (see Fig. 1 legend) might present a formidable obstacle to both messenger RNA chain initiation and elongation. This has become an important issue, as evidence has accumulated that core histone octamers and some histone H1 remain bound to the DNA of transcriptionally active genes¹⁻³. However, the 30-nm fibre in the neighbourhood of these genes is at least partially disrupted^{4,5}, probably through depletion of histone H1 (or H5) (refs 6, 7), which binds both to the core histone/DNA complex and to linker DNA, and normally stabilizes the fibre.

Inhibitory action of histones

The presence of bound core histone octamers is not necessarily detrimental to transcription. Experiments *in vitro* with chromatin templates show that RNA polymerases, once started, are able to transcribe through core histone octamers associated with the body of the gene. Chain elongation by SP6 polymerase and RNA polymerase II is unimpeded by the presence of one or two histone octamers in their path^{8,9}. When longer arrays of octamers are present, pol II is still able to elongate RNA *in vitro*, but transcription is partially inhibited by pausing along the template¹⁰, an effect that may well be alleviated *in vivo* by histone modification or the presence of additional factors. The mechanisms by which the polymerase is able to transcribe through nucleosome-covered DNA *in vitro* or *in vivo* are not well understood¹¹, but might involve either direct transfer of octamers or their components out of the path of the transcription complex to another part of the template¹², unfolding of the octamer¹³, or disruption of the octamer with a change in the histone domains that would interact with DNA¹⁴.

Although the chain elongation process may not be drastically impeded by nucleosomes, initiation of transcription *in vitro* (see Fig. 2 legend for a summary of mechanisms thought to regulate initiation) is inhibited if the histone octamer is bound over the promoter^{8,9,15}. This suggests simple models in which the transcription complex (and *trans*-acting factors that can bind nearby) compete with histones for sites on the promoter. One possibility is a dynamic quasi-equilibrium in which transcription factors and histone octamers compete for binding sites in the non-dividing cell; evidence for the existence of such mechanisms will be described below. In other cases the binding might be functionally irreversible, so that the first components to bind win the competition, leading either to activation or repression of the gene. Consistent with this mechanism, chromatin assembly experiments *in vitro* have shown that tissue-specific factors bound near a β -globin promoter are able to exclude nucleosomes, but that this does not occur if the histones are added first¹⁶. This 'pre-emptive' competition was originally invoked to explain the regulation of expression of the gene for 5S RNA by

the *trans*-acting factors associated with RNA polymerase III transcription¹⁷⁻¹⁹.

Several groups^{15,20-22} have examined the effect on transcription of competitive binding of nucleosomes and pol II transcription factors at the promoter. In typical experiments, *Xenopus* egg or oocyte extracts were used to assemble nucleosomes onto a DNA template, either before, during or after the addition of transcription factors. Knezetic *et al.*²³ used HeLa cell nuclear extracts to form preinitiation transcription complexes on an adenovirus type-2 major late promoter (MLP). Consistent with the pre-emptive model, transcription was observed if the

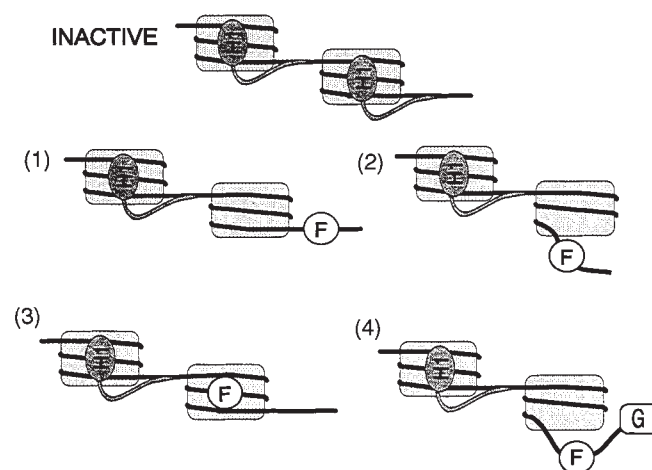


FIG. 1 Possible mechanisms of direct disruption of nucleosomes by *trans*-acting factors. Histone octamers are shown as grey rectangles. They are composed of two molecules of each of the 'core' histones: H2A, H2B, H3 and H4. Two superhelical turns of DNA containing a total of 165 base pairs (bp) of DNA are wrapped around the octamer. A molecule of a fifth histone, H1, is bound to the exterior of each subunit at the point where the DNA exits. The C-terminal tail of histone H1 interacts with DNA in the linker region; the length of linker DNA can vary from 0 to about 80 bp. The polynucleosome filament is further compacted *in vivo* by coiling to form a solenoidal structure with about six nucleosomes per turn. The resulting fibre is about 30 nm in diameter (see ref. 85 for an extensive description of chromatin structure). Higher levels of compaction are also observed within the nucleus, but are less well understood. Little mention is made in this review of the nature or function of 'constitutive' heterochromatin found at centromeric and telomeric sequences, or the position-effect variegation associated with such sequences (see ref. 84 for review). The structures shown in this figure are numbered in correspondence with their description in the text. Inactive (top): an unperturbed polynucleosome filament with core histone octamers and histone H1 in place. Possible roles of *trans*-acting factors: (1) factor F binds in the linker and displaces H1 from the adjacent octamer. Binding of F to the linker would presumably displace the H1 C-terminal tail, and greatly reduce the affinity of H1 for the nucleosome, but H1 could in principle still remain bound through its globular domain; (2) factor F binds to DNA just inside the place where it enters the nucleosome; (3) factor F binds on the dyad axis; (4) factor G disrupts nucleosome structure and factor F binds. The histone octamer may unfold, slide or completely dissociate as a result of some of these reactions.

preinitiation complex was formed before, but not after, chromatin assembly. Similar results were obtained by Workman and Roeder²¹, who showed further that a fraction enriched in TFIID was sufficient to protect the activity of the promoter against subsequent nucleosome binding; cloned yeast TFIID has also been used²².

In these experiments, transcription from the adeno-2 MLP does not occur if TFIID and histones are added simultaneously (rather than sequentially). However, the presence of *trans*-acting factors can affect this behaviour. The addition of USF (an upstream transcription factor), which binds at position -58 of the adeno-2 MLP, makes it possible to assemble a transcriptionally active complex on the promoter in the simultaneous presence of both TFIID and histones^{24,25}. These experiments are more difficult to interpret than those in which histones and TFIID are added sequentially. When transcription factors and histones are mixed in the assembly reaction, the winner of the race to the promoter may be determined by kinetic processes peculiar to the reaction conditions. For example, Wu²⁶ has found in the case of the *Drosophila hsp70* promoter that if TFIID and histones are present simultaneously in the chromatin assembly reaction, a 'potentiated' chromatin template is produced, capable of responding to subsequent addition of heat-shock transcription factor. In this case no other transcription factors are required during chromatin assembly, suggesting that a modest change in conditions or in the TFIID binding site may be sufficient to alter the experimental results.

Similar methods have been used to explore the possible role of other *trans*-acting factors in the assembly process. Workman *et al.*²⁷ have investigated the effect of transactivating GAL4 derivatives (GAL4-VP16, GAL4-AH) on a plasmid containing five tandem repeats of a GAL4 binding site adjacent to an *hsp70* promoter (reviewed in ref. 28). The presence of GAL4-VP16 or GAL4-AH during chromatin assembly led to formation of a transcriptionally active template. In this case it was not necessary that the basal factors (including TFIID) be present during the assembly process: they could be added afterwards. The transactivating domains of the GAL4 derivatives were essential. GAL4 (1-94), which lacks these domains, was unable to prevent inactivation by histones, despite the fact that it was fully capable of binding to all five GAL4 sites. It may be that the domains contribute to the disruption of adjacent nucleosomes.

In assessing all of these competition results, it must be kept in mind that the inclusion of transcription factors during chromatin assembly has never been observed to produce a significant stimulation of transcription relative to basal levels on the corresponding histone-free plasmid. The observed 'stimulation' is actually a relief of repression, measured relative to the histone-covered template assembled in the absence of factors. It may be that for some genes the principal role of *trans*-acting factors *in vivo* is to allow basal level transcription, but it is not likely to be their role for all genes. A further complication arises because in all of the experiments *in vitro* only a small fraction of the templates is actually transcribed. In the only report to address this question²³, it was estimated that 3% of the DNA was transcribed, but it was demonstrated that the transcribed template was in fact packaged with nucleosomes. The possibility still exists in most of these studies that the active templates differ in significant ways from the preponderant species.

Although transcription experiments *in vitro* are instructive for working out the possible interactions of histones and transcription factors, the constructions and the experimental conditions are often quite different from those found within the cell; the models must ultimately be tested by examination of individual genes in their natural setting. This requires mapping of nucleosome positions around the promoter *in vivo* in cells in which the gene is switched on or off, identification of the factors that actually bind there, and the eventual reconstruction of the system *in vitro*. A promising new direction comes from the use of systems that replicate DNA *in vitro*; competition between transcription

complex assembly and chromatin assembly can be studied under more realistic conditions in these systems, although the method has been used so far only for RNA polymerase III-dependent transcription²⁹. In a few instances, it has been possible to demonstrate a dependence of transcriptional activation on replication³⁰.

It seems reasonable to suppose that in these experiments core histones are absent from promoters that have been occupied successfully by transcription factors. As the templates transcribed *in vitro* represent such a small fraction of the total, it is difficult to verify this directly, although in the case of activation by USF a restriction endonuclease probe was used to show that some kind of nucleosome displacement had occurred in 51% of the copies²⁴. The situation *in vivo* is clearer. Promoters and enhancers are typically marked by hypersensitive domains in chromatin that reflect the absence of a canonical nucleosome (see refs 1 and 31 for reviews). The assessment of structure is limited by the nature of the probes, and it is therefore not clear in any given case whether the histone octamers are completely displaced (by sliding or dissociation), or whether they remain attached at the promoter in some cryptic form that cannot be detected in footprinting experiments.

Displacement of histones by activators

It can be argued that a purely pre-emptive mechanism relegates the histones to a role of mere nuisance. *Trans*-acting factors need only get to their sites first and hold on, and the histones can then be ignored. Recent results in yeast (*S. cerevisiae*) make it clear that the histones can also play a well defined biochemical role, in which they interact, directly or indirectly, with other factors through defined domains to alter nucleosome position and the level of expression of specific genes.

The effect of *trans*-acting factors on nucleosome positioning is demonstrated dramatically in the behaviour of the *PHO5* gene promoter. *PHO5* is an acid phosphatase gene that is induced in the absence of phosphate. Hörz and his co-workers^{32,33} have shown that when the gene is induced a cluster of four positioned nucleosomes at the promoter is removed or disrupted. In the repressed state, one of these nucleosomes (nuc-2) covers a pair of binding sites for the *trans*-activating proteins PHO2 and PHO4; a second PHO4 site lies in the linker DNA between that nucleosome and its 5' neighbour (nuc-3). Analysis of mutants shows that *PHO2* and *PHO4* are essential for the generation of the active chromatin structure³⁴, as well as for the expression of *PHO5* (ref. 35). Further experiments suggest a series of events during induction in which PHO4 binds to the site within the linker and with the assistance of PHO2 displaces the adjacent nucleosome (nuc-2), exposing the second PHO4 site. *PHO2* can in some circumstances be dispensed with; in a *pho2* null mutant, overexpression of *PHO4* generates an active chromatin structure. It is not clear how the complex then disrupts or displaces the other three nucleosomes. The energetic balance that results in chromatin disruption is in any case a delicate one; changes in the DNA sequence under nuc-2 can alter the susceptibility of the promoter both to induction and chromatin disruption³⁶.

The interplay between DNA sequence, the binding of *trans*-acting factors, and nucleosome placement is also well illustrated by recent work on the effect of the yeast $\alpha 2$ repressor on chromatin structure. These experiments³⁷ made use of an autonomously replicating yeast TRP1/ARS1 plasmid, into which an $\alpha 2$ operator (which contains binding sites for the proteins $\alpha 2$ and MCM1) had been introduced near a nuclease-hypersensitive region upstream of the *TRP1* gene. In α cells, which express $\alpha 2$, *TRP1* expression was reduced, nucleosomes were positioned precisely over sequences flanking the operator, the plasmid acquired an additional nucleosome, and the hypersensitive site disappeared; in α cells, which do not express $\alpha 2$, no exact positioning was observed. The results suggest that the binding of $\alpha 2$ -MCM1 may act on expression by fixing the positions of critical nucleosomes on or near the gene.

Nucleosomes obviously can affect transcription by blocking critical DNA sequences. In some cases, however, the DNA bending induced by a nucleosome may be used to bring together distant DNA sequence elements, which can facilitate the interaction and binding of *trans*-acting factors. Such a mechanism has been proposed to explain footprinting data obtained at the *Drosophila hsp26* heat-shock gene promoter³¹.

Effects of histone mutations

It is clear from studies of yeast histone mutants that the effects of histones on transcription reflect the operation of an elaborate and precise mechanism. Because the yeast genome contains only two copies of each core histone gene and only one copy is required, the consequences of gene replacement can be examined. Grunstein and his collaborators³⁸ found that deletions in the N terminus of histone H4 lead to transcriptional derepression of the yeast silent mating loci, *HML α* and *HMRA*. Replacement of any of four basic amino acids in the H4 tail (Lys 16, Arg 17, His 18, or Arg 19) with glycine was sufficient to reduce mating efficiency over 5,000-fold³⁹⁻⁴¹. The Gly-16, Gln-16, Gly-17 or Gly-18 mutants were all suppressed by either of two mutations in *SIR3*, a protein required for repression of *HML α* and *HMRA*. Each suppressor *SIR3* mutant contains a single amino-acid change in the N terminus of the protein.

The N-terminal histone tails, with their many positive charges, have been implicated in the condensation of chromatin into higher-order structures⁴². No histone H1-like protein has yet been identified in yeast (but see ref. 43), but it is possible that some internucleosomal packing is mediated by core histone tails^{44,45}, and that this affects nucleosome position and stability. It is also possible that the interaction of tails with DNA affects the ability of the octamer to slide along DNA. It should be recalled that the reversibly acetylated lysine residues of the H4 tail are located at positions 5, 8, 12 and 16; there is considerable evidence that transcriptionally active chromatin carries hyperacetylated histones.

Histone H4 tails are involved not only in repression, but in gene activation as well. Deletion of residues 4-23, or substitution by glutamine of all four lysines (5, 8, 12 and 16) results in an order of magnitude reduction in *GAL1* induction by galactose⁴⁶. It has been known for some time⁴⁷ that *PHO5* is activated in yeast mutants in which histone H4 synthesis is shut off, and in which nucleosomes are depleted. The induction of *PHO5* is also affected by mutations in the H4 tails, although to a lesser extent than is *GAL1*. The nucleosome displacement phenomena already described for *PHO5* are thus strongly linked to the properties of the histone H4 tails. So too are the positioning effects of the $\alpha 2$ repressor, which depend upon these tails (S. Roth *et al.*, personal communication). But silent mating locus repression and *GAL1/PHO5* activation are affected quite differently by individual amino-acid replacements, suggesting that the complexities of histone H4 function are by no means yet revealed. Not all genes are affected by H4 mutations; constitutive genes that have been examined, and which may not have nucleosomes positioned on their promoters, are not affected. None of these effects is seen with mutants of the other core histones, but in the case of H3 there is some evidence for a different set of functions⁴⁶. Further experiments with mutants in which H4 synthesis is turned off show that depletion of nucleosomes results in nearly full induction of expression for some genes (*CUP1*, *HIS3*). This suggests that nucleosome placement may be a major element in regulation of such genes (M. Grunstein, personal communication).

How activators invade nucleosomes

These regulatory processes in yeast do not appear to involve a pre-emptive mechanism of chromatin transcriptional activation. In the case of *PHO5* induction, it has been shown (W. H  rz, personal communication) that DNA replication is not required for the displacement of octamers near the promoter, and there

is no evidence so far that replication is required for any of the other chromatin rearrangements in yeast described above (although it is known that silencing at the silent mating type loci requires replication⁴⁸). In these cases, *trans*-acting factors must be capable of disrupting or displacing existing nucleosomes, which implies that the factors must bind to DNA on or near the histone octamer (see below). The induction of *PHO5* discussed above employs such a mechanism, and a related mechanism has been described in some detail for the glucocorticoid receptor⁴⁹⁻⁵¹.

A recent paper that may be relevant to this question explores the binding of GAL4-VP16 to DNA sites that are complexed in nucleosomes⁵². Work from the same group, discussed above, has established conditions in which pre-emptive binding occurs *in vitro*. The new data show that GAL4 and its derivatives at higher concentrations can also bind to DNA in chromatin, with an estimated affinity that is 100-fold lower (for a single site) or 10-fold lower (for five sites in tandem) than binding to the corresponding naked DNA template. Gel electrophoresis reveals that DNA, GAL4 and histones form a complex of undetermined structure. Surprisingly, the rotational position of the single site on the nucleosome surface does not affect the ability to bind. In contrast to the behaviour of GAL4, human heat shock factor is unable to bind to a site which is complexed in a nucleosome. These results suggest that some *trans*-acting factors may invade nucleosomes. It is not clear whether GAL4 itself ever functions this way *in vivo*, given its reduced affinity for nucleosomal DNA.

The issue of histone displacement is further complicated by the participation of histone H1 in stabilization of nucleosomes and formation of higher order structure. The possible role of histone H1 in regulation has been investigated⁵³ in a series of

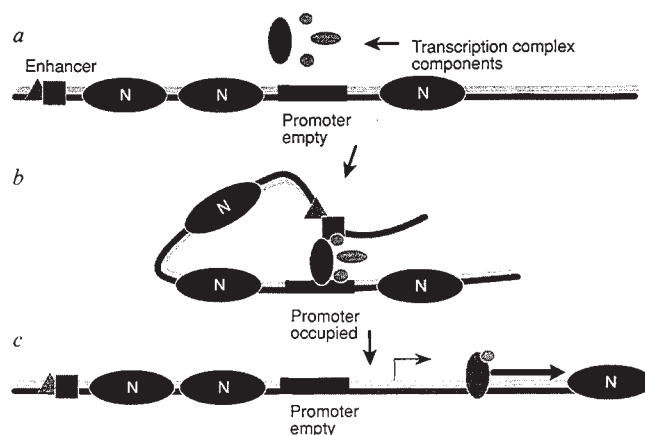


FIG. 2 Activation of transcription by the interaction of DNA-bound *trans*-acting factors and target proteins in the transcription complex. According to proposed models^{62,63} such factors function by making contacts with certain target components of the transcription complex bound near the 5' end of the gene (perhaps TFIID but, at least in some cases, more probably TFIIB (ref. 28)), in such a way as to alter the transcription initiation rate. Many *trans*-activating factors carry both the sequence-specific DNA binding site and a *trans*-activating domain that makes contact with the target directly. Other factors may bind a second protein, which in turn contacts the target. Activation may involve stabilization of DNA binding of the target protein and/or induction of changes in its conformation. Note that transcriptional activation need not involve cooperative interaction between *trans*-acting factors to increase binding of these factors, but in some cases does involve an additive, synergistic effect of these factors on the transcription complex^{66,67}. But if the *trans*-acting factors are partially dissociated from their binding sites, favourable interactions between the factors will lead to increased site occupancy, and presumably higher levels of activation. Transcription factors are transiently stabilized (b) at the promoter by interaction with other components bound at the enhancer. Though the promoter may be occupied by transcription factors much of the time, stimulation of transcription does not require that the promoter always be occupied (c). Nucleosomes are marked N.

experiments similar in design to those described above which measured the competition of core histones and transactivators for sites on DNA. In such experiments, binding of H1 to DNA (in the absence of core histones) inhibited transcription of a test gene containing flanking binding sites for GAL4-VP16, Sp1 or the GAGA factor. Addition of the appropriate factor relieved this inhibition. Histone H1 binds to DNA less tightly than do the core histones, and so it is not surprising that order of addition is less important: GAL4-VP16, even if added after H1, was able to reverse the inhibition. These studies have recently been extended⁵⁴ by adding H1 to DNA templates previously reconstituted with core histones. The effects of H1 on transcription were in many ways similar to those observed in the earlier experiments with naked DNA. As with all the *in vitro* transcription experiments described above, it is speculative to draw conclusions about mechanisms *in vivo*. Nonetheless, the results⁵⁴ demonstrate, in a single system, irreversible repression by core histones, semi-reversible inhibition by H1, and activation above the basal level by GAL4-VP16 on chromatin templates.

Many of the transcription experiments carried out with RNA polymerase II have their antecedents in studies of RNA polymerase III transcription on chromatin templates. In particular, the earliest evidence of a regulatory role for histone H1 comes from studies some years ago of 5S ribosomal gene transcription by pol III (ref. 55). The results suggested that chromatin templates carrying the core histones but lacking H1 could be transcribed, and that the addition of H1 prevented the formation of preinitiation complexes. More recent results (reviewed in refs 19 and 56) show that addition of H1 selectively represses certain class III genes.

Dynamic versus pre-emptive competition

In summary, transcriptional activation experiments suggest that there are two kinds of mechanisms for disruption of chromatin structure. The first mechanism, dynamic competition (Fig. 1), does not require DNA replication to make regulatory regions accessible to transactivators, for one or more of the following reasons; (1) transactivator binding sites may be located in the linker DNA between positioned nucleosomes, so that it is a relatively easy task to get at them by displacing histone H1; (2) the sites may be relatively accessible if they are within the nucleosome but situated near the DNA entrance or exit points⁵⁷; (3) the DNA-binding site, though located near the pseudodyad of the nucleosome (that is, about one superhelical turn away from the entrance), may still be accessible on the surface if the DNA recognition sequence is sufficiently short and correctly positioned; (4) even though the binding site is within the nucleosome and inaccessible, another *trans*-acting factor that binds nearby may be able to destabilize or displace the histone octamer. Displacement might involve sliding, jumping, or complete dissociation from the DNA. A factor has been described which, as its principal function, appears to generate a nucleosome-free region⁵⁸, but it is unknown whether it operates by displacement or by the mechanism discussed later.

The second mechanism is the pre-emptive competition model described earlier. For this class of promoters and enhancers, the core histones block the factor binding sites in such a way that the sites are inaccessible to transactivators in the resting cell—perhaps because of the position of octamers relative to the binding sites, or because the region, if completely covered by octamers, is able to fold rapidly into a compact higher-order structure. Under these circumstances, chromatin changes would presumably occur only during replication, when the site is at least partially exposed at the replication fork before chromatin assembly. It is assumed that a nucleosome-free region, once established, is maintained in resting cells because of the absence of large pools of unbound histones, and of active assembly mechanisms. Before elaborating on this model, it is important to qualify the concept of 'irreversible' inactivation of chromatin. As used here, it concerns only the putative resistance of certain

inactive chromatin regions to activation in the absence of DNA replication. There is no question that many kinds of dividing differentiated cells can be reprogrammed⁵⁹. We also know from cell fusion studies⁶⁰ that formation of stable heterokaryons can lead to reprogramming of gene expression without DNA replication. Similarly, Baron and Maniatis⁶¹ have studied gene expression in transient heterokaryons formed by fusion of adult mouse erythroleukaemia (MEL) cells and human embryonic erythroid (K562) cells with a number of types of mouse or human nonerythroid cells. Such fusions, in which the nuclei remain separate and distinct, result in activation of expression of many stage-specific members of the globin gene family in the nonerythroid nucleus. In these experiments, *trans*-acting factors from the erythroid cell nucleus are capable of exerting a dominant effect on expression within the non-erythroid nucleus. The events of cell fusion may well labilize histones in a way that makes protected DNA sites accessible. It is therefore not clear from these results whether a normal, non-dividing cell could also be reprogrammed by the expression of heterospecific factors. It is also not known whether the formation of heterokaryons can induce the long-range changes in chromatin structure (see below) associated with normal activation of gene clusters.

It seems likely that some regulatory regions in nondividing cells are packaged in chromatin structures that are resistant to direct nucleosome displacement by *trans*-acting factors. These structures could only be disrupted by the pre-emptive mechanism. Here the transcription experiments *in vitro* described above must serve as an admittedly imperfect model. What is required of a transactivation complex to survive the passage of the replication-dependent nucleosome assembly mechanism is different from what is required for stimulation of transcription. Transcriptional activation merely demands that the transactivator raise the average level of occupancy of the promoter by the rate-determining transcription factors (Fig. 2; refs 62, 63). This can be accomplished even if the sites are not occupied all the time. In contrast, reliable creation of an altered chromatin state by pre-emption can only occur if the promoter-bound complex is completely stable during the critical time period when nucleosomes (or some crucial subset of core histones) are assembled.

Such stabilization can of course be provided by the kinds of intermolecular contacts that are thought to occur in enhancer-promoter interactions. Many *trans*-acting factors bound to DNA can interact with each other to form homodimers or heterodimers. For example, two molecules of progesterone receptor⁶⁴ or Sp1 (ref. 65), or one molecule each of Sp1 and the BPV enhancer E2 protein⁶⁶, bound at widely separated sites, induce loop formation in DNA. The favourable free energy of this dimerization assures that binding of one of these factors at an enhancer will stabilize its mates bound near the promoter. Under suitable conditions, the occupancy of promoter sites will be increased.

Role of locus control regions

The properties of locus control regions (LCRs, formerly called dominant control (DCR) or locus activating (LAR) regions) suggest that they may play a special part in the stabilization of open chromatin. Locus control regions are *cis*-acting elements that make the expression of associated genes or gene clusters independent of their position within the genome⁶⁷⁻⁷⁰. In the chromatin of the human β -globin gene cluster, the LCR is marked by a set of four strongly nuclease-sensitive domains located 10–20 kilobases upstream of the genes^{67,68,71}. If this LCR is coupled to a human β -globin gene (with its promoter and other local regulatory elements) and introduced into transgenic mice, the amount of human β -globin mRNA synthesized is proportional to the number of gene copies that have been integrated. The proportionality suggests that the LCR allows gene expression independent of the chromatin environment at the site of integration; without the LCR, integration of the β -globin gene typically leads to low levels of expression that

are erratically related to copy number. The full human β -globin LCR also confers the ability to express the gene at high levels in transgenic mice. Although this is an enhancer-like effect, it is in most cases distinguishable from canonical enhancer activity: Three of the four human β -globin LCR domains do not function as enhancers in transient expression assays, but only when stably integrated, suggesting that the activity is associated with chromatin structure⁷²⁻⁷⁵. This stimulatory behaviour and the property of position independence are partially separable. In the case of the chicken β -globin cluster, an element has been identified that confers position independence without high levels of expression⁷⁶; similar results are obtained with mutated versions of some of the individual domains of the human β -globin LCR (refs 75, 77, 78).

These LCR domains are full of DNA-binding sequences for well-known *trans*-acting factors. The domains look as though they were put together by an overenthusiastic student determined to construct a powerful *cis*-acting element. For example, the human β locus domain 5'HS2 contains within a 300-base pair segment four binding sites for GATA-1, and a site for NFE-2, as well as three binding sites for fully or partially erythroid-specific factors and two sites for ubiquitous factors⁷³. It is possible that these LCR domains have evolved from enhancers, but they may have a different function. It was pointed out above that the requirements for preventing nucleosome assembly during replication are more stringent than those required for transcriptional activation. I suggest that a primary role of the LCR is to enter into cooperative binding interactions that keep the promoter free of histones. (A similar role has been proposed by Wolffe for enhancers⁷⁹.) In this view, the reason for the 'overkill', implicit in the multiple factor sites within the LCR, is to guarantee that the promoter-bound *trans*-acting factors stay in place during replication (Fig. 3). Different LCR domains may be targeted to different promoters (which are themselves characterized by different locally bound factors). Of course a byproduct of such a design would often be enhancer activity—not surprising as some of the same mechanisms are involved, but perhaps with a different purpose here.

The effect of this activity would be to confer position-independence in transgenic experiments, because the target promoter elements alone would seldom become functional when integrated into the genome. This position-independent effect can be distinguished from the buffering or insulating effect generated by the *scs* elements of *Drosophila* characterized by Kellum and Schedl⁸⁰, which must be placed on both sides of a gene and its local control elements to assure position-independent expression. The 'A elements' near the chicken lysozyme gene⁸¹ might function similarly; it is not clear how these are related functionally to the scaffold-associated regions (SARs) described by Laemmli^{82,83}. All of these buffering elements may work by preventing both the propagation of an inactive chromatin structure from neighbouring parts of the genome, and the extension of an active chromatin conformation into adjacent inactive chromatin (although they may not be sufficient to insulate against the effects of integration into centromeric or other constitutive heterochromatin^{80,84}). In contrast, the LCR activity I propose above would provide a dominant signal that overrides the inhibition. Evidence at least consistent with such a model comes from the observation (M. Reitman, E. Lee, H. Westphal and G.F., unpublished results) that a globin LCR element alone can be hypersensitive in erythroid cells of transgenic mice, but that a globin promoter is not hypersensitive except in the presence of an LCR element.

It should be clear that histones play active and varied roles in transcription by RNA polymerase II. In those instances in which a dynamic competition exists between histones and *trans*-

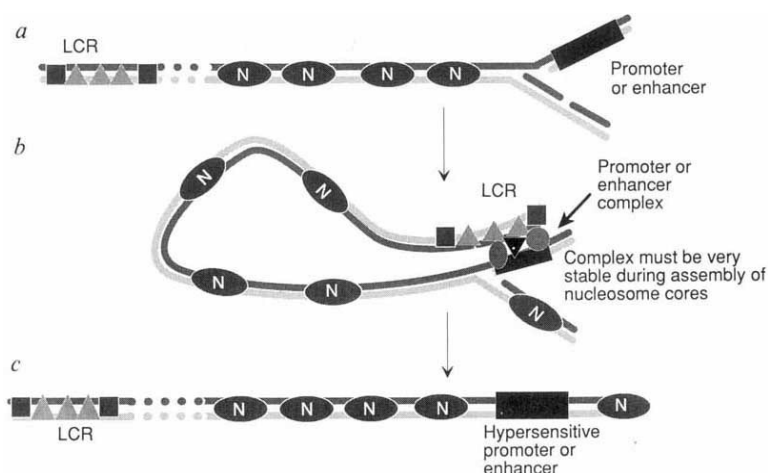


FIG. 3 Proposed model for LCR action. Transcription factor binding is stabilized at the replication fork by the action of factors bound at the LCR. Unlike the situation in Fig. 2, the complex must not dissociate during chromatin assembly if a hypersensitive site is to be generated.

acting factors, the conserved features of histone sequence are likely to be important in this function; some factors are designed to take advantage of those features. Nucleosomes are not merely compact and featureless inducers of DNA bending; they are biochemically active complexes. Chromatin structure at the level of the individual nucleosome, the 30-nm fibre, and the loop domain has an intimate connection with the transcriptional machinery, and we will have to learn much more about chromatin to understand the eukaryotic transcription process. □

Gary Felsenfeld is at the Laboratory of Molecular Biology, National Institutes of Health, Bethesda, Maryland 20892, USA.

- Gross, D. S. & Garrard, W. T. *Rev. Biochem.* **57**, 159-197 (1988).
- Grunstein, M. *Rev. Cell Biol.* **6**, 643-678 (1990).
- Erickson, C., Grossbach, U., Björkroth, B. & Daneholt, B. *Cell* **60**, 73-83 (1990).
- Kimura, T., Mills, F., Allen, J. & Gould, H. *Nature* **306**, 709-712 (1983).
- Fisher, E. A. & Felsenfeld, G. *Biochemistry* **25**, 8010-8016 (1986).
- Postnikov, Y. V. et al. *Nucleic Acids Res.* **19**, 717-725 (1991).
- Kamakaka, R. T. & Thomas, J. O. *EMBO J.* **9**, 3997-4006 (1990).
- Losa, R. & Brown, D. D. *Cell* **50**, 801-808 (1987).
- Lorch, Y., LaPointe, J. W. & Kornberg, R. D. *Cell* **49**, 203-210 (1987).
- Izban, M. G. & Luse, D. S. *Genes Dev.* **5**, 683-696 (1991).
- Thoma, F. *Trends Genet.* **7**, 175-177 (1991).
- Clark, D. J. & Felsenfeld, G. *EMBO J.* **10**, 387-395 (1991).
- Lee, M.-S. & Garrard, W. T. *EMBO J.* **10**, 607-615 (1991).
- Nacheva, G. et al. *Cell* **58**, 27-36 (1989).
- Knezetic, J. A. & Luse, D. S. *Cell* **45**, 95-104 (1986).
- Emerson, B. M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 95-99 (1984).
- Bogenhagen, D. F., Wormington, W. M. & Brown, D. D. *Cell* **28**, 413-421 (1982).
- Gottesfeld, J. & Bloomer, L. S. *Cell* **28**, 781-791 (1982).
- Wolffe, A. P. & Brown, D. D. *Science* **241**, 1626-1632 (1988).
- Matsui, T. *Molec. cell. Biol.* **7**, 1401-1408 (1987).
- Workman, J. L. & Roeder, R. G. *Cell* **51**, 613-622 (1987).
- Meisterernst, M., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9153-9157 (1990).
- Knezetic, J. A., Jacob, G. A. & Luse, D. S. *Molec. cell. Biol.* **8**, 3114-3121 (1988).
- Workman, J. L., Roeder, R. G. & Kingston, R. E. *EMBO J.* **9**, 1299-1308 (1990).
- Workman, J. L., Abmayr, S. M., Cromlish, W. A. & Roeder, R. G. *Cell* **55**, 211-219 (1988).
- Becker, P. B., Rabindran, S. K. & Wu, C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4109-4113 (1991).
- Workman, J. L., Taylor, I. C. A. & Kingston, R. E. *Cell* **64**, 533-544 (1991).
- Sharp, P. A. *Nature* **351**, 16-18 (1991).
- Almouzni, G., Mechali, M. & Wolffe, A. P. *EMBO J.* **9**, 573-582 (1990).
- Enver, T., Brewer, A. C. & Patient, R. K. *Molec. cell. Biol.* **8**, 1301-1308 (1988).
- Elgin, S. C. R. *J. biol. Chem.* **263**, 19259-19262 (1988).
- Almer, A. & Hörz, W. *EMBO J.* **5**, 2681-2687 (1986).
- Almer, A., Rudolph, H., Hinnen, A. & Hörz, W. *EMBO J.* **5**, 2689-2696 (1986).
- Fascher, K.-D., Schmitz, J. & Hörz, W. *EMBO J.* **9**, 2523-2528 (1990).
- Yoshida, K., Ogawa, N. & Oshima, Y. *Molec. gen. Genet.* **217**, 40-46, (1989).
- Straka, C. & Hörz, W. *EMBO J.* **10**, 361-368 (1991).
- Roth, S. Y., Dean, A. & Simpson, R. T. *Molec. cell. Biol.* **10**, 2247-2260 (1990).
- Kayne, P. S. et al. *Cell* **55**, 27-39 (1988).
- Johnson, L. M., Kayne, P. S., Kahn, E. S. & Grunstein, M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6286-6290 (1990).
- Megee, P. C., Morgan, B. A., Mittman, B. A. & Smith, M. M. *Science* **247**, 841-845 (1990).
- Park, E. C. & Szostak, J. W. *Molec. cell. Biol.* **10**, 4932-4934 (1990).
- Allan, J., Harborne, N., Rau, D. C. & Gould, H. *J. Cell Biol.* **93**, 285-297 (1982).
- Lowary, P. T. & Widom, J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8266-8270 (1989).
- McGhee, J. D., Rau, D. C., Charney, E. & Felsenfeld, G. *Cell* **22**, 87-96 (1980).

45. McGhee, J. D. & Felsenfeld, G. A. *Rev. Biochem.* **49**, 1115–1156 (1980).
46. Durrin, L. K., Mann, R. K., Kayne, P. S. & Grunstein, M. *Cell* **65**, 1023–1031 (1991).
47. Han, M. & Grunstein, M. *Cell* **55**, 1137–1145 (1988).
48. Miller, A. M. & Nasmyth, K. A. *Nature* **312**, 247–251 (1984).
49. Cordingley, M. G., Riegel, A. T. & Hager, G. L. *Cell* **48**, 261–270 (1987).
50. Perlman, T. & Wrangé, Ö. *EMBO J.* **7**, 3073–3079 (1988).
51. Piña, B., Brüggemeier, U. & Beato, M. *Cell* **60**, 719–731 (1989).
52. Taylor, I. C. A., Workman, J. L., Schuetz, T. J. & Kingston, R. E. *Genes Dev.* **5**, 1285–1298 (1991).
53. Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. & Kadonaga, J. T. *Science* **251**, 643–649 (1991).
54. Laybourn, P. J. & Kadonaga, J. T. *Science* **254**, 238–245 (1991).
55. Schlissel, M. S. & Brown, D. D. *Cell* **37**, 903–913 (1984).
56. Wolffe, A. P. *New Biologist* **2**, 211–218 (1990).
57. Simpson, R. T. *Nature* **343**, 387–389 (1990).
58. Chasman, D. I. et al. *Genes Dev.* **4**, 503–514 (1990).
59. Blau, H. M. & Baltimore, D. *J. Cell Biol.* **112**, 781–783 (1991).
60. Blau, H. M., Chiu, C.-P. & Webster, C. *Cell* **32**, 1171–1180 (1983).
61. Baron, M. H. & Maniatis, T. *Cell* **46**, 591–602 (1986).
62. Ptashne, M. *Nature* **335**, 683–689 (1988).
63. Ptashne, M. & Gann, A. *Nature* **346**, 329–331 (1990).
64. Theveny, B. et al. *Nature* **329**, 79–81 (1987).
65. Su, W., Jackson, S., Tjian, R. & Echols, H. *Genes Dev.* **5**, 820–826 (1991).
66. Li, R., Knight, J. D., Jackson, S. P., Tjian, R. & Botchan, M. R. *Cell* **65**, 493–505 (1991).
67. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. *Cell* **51**, 975–985 (1987).
68. Forrester, W. C., Takegama, S., Papayannopoulou, T., Stamatiyannopoulos, G. & Groudine, M. *Nucleic Acids Res.* **15**, 10159–10177 (1987).
69. Evans, T., Felsenfeld, G. & Reitman, M. A. *Rev. Cell Bio.* **6**, 95–124 (1990).
70. Townes, T. & Behringer, R. R. *Trends Genet.* **6**, 219–223 (1990).
71. Tuan, D., Solomon, W., Li, Q. & London, I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6384–6388 (1985).
72. Tuan, D., Y. H., Solomon, W. B., London, I. M. & Lee, D. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2554–2558 (1989).
73. Talbot, D., Philipsen, S., Fraser, P. & Grosveld, F. *EMBO J.* **9**, 2169–2178 (1990).
74. Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. *EMBO J.* **9**, 2159–2167 (1990).
75. Talbot, D. & Grosveld, F. *EMBO J.* **10**, 1391–1398 (1991).
76. Reitman, M., Lee, E., Westphal, H. & Felsenfeld, G. *Nature* **348**, 749–752 (1990).
77. Fraser, P., Hurst, J., Collis, P. & Grosveld, F. *Nucleic Acids Res.* **18**, 3503–3508 (1990).
78. Caterina, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1626–1630 (1991).
79. Wolffe, A. P. in *Molecular Endocrinology and Steroid Hormone Action* 171–186 (Liss, New York, 1990).
80. Kellum, R. & Schedl, P. *Cell* **64**, 941–950 (1991).
81. Stief, A., Winter, D. M., Stratling, W. H. & Sippel, A. E. *Nature* **341**, 343–345 (1989).
82. Gasser, S. M. & Laemmli, U. K. *Trends Genet.* **3**, 16–21 (1987).
83. Izaurralde, E., Käs, E. & Laemmli, U. K. *J. molec. Biol.* **210**, 573–585 (1989).
84. Wilson, C., Bellen, H. J. & Gehring, W. J. A. *Rev. Cell Biol.* **6**, 679–714 (1990).
85. van Holde, K. E. *Chromatin* (Springer, New York, 1989).
86. Lin, Y.-S., Carey, M., Ptashne, M. & Green, M. R. *Nature* **345**, 359–361 (1990).
87. Carey, M., Lin, Y. S., Green, M. R. & Ptashne, M. *Nature* **345**, 361–364 (1990).

ACKNOWLEDGEMENTS. I am grateful to Bruce Alberts, Michael Grunstein, Wolfram Hörz, Marc Reitman, Robert Simpson, and Carl Wu and to the members of the Laboratory of Molecular Biology, NIDDK, for their instructive criticism, and for the opportunity to cite unpublished data and read manuscripts before publication.

ARTICLES

Mapping T-cell receptor–peptide contacts by variant peptide immunization of single-chain transgenics

Jeffrey L. Jorgensen, Ursula Esser, Barbara Fazekas de St. Groth^{*†}, Philip A. Reay & Mark M. Davis^{*}

Department of Microbiology and Immunology and ^{*}The Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA

To test models of T-cell recognition, mice transgenic for T-cell receptor α or β chain have been immunized with variant peptides that force changes in the resulting T-cell response. In particular, charge substitutions on the peptide often elicit reciprocal charges in the junctional (CDR3) sequences of T-cell receptor V_α or V_β chains, indicating direct T-cell receptor–peptide contact, and allowing derivation of a topology for the T-cell receptor–MHC interaction. At one position on the peptide, variants transformed a homogeneous V_β response into a very heterogeneous one.

THE interaction of T-cell receptors (TCRs) with their ligands, peptide fragments bound to major histocompatibility complex (MHC) molecules, is central to the initiation and propagation of most immune responses. TCR heterodimers arise from similar types of V, D and J rearranging gene elements as immunoglobulins, and share many of the conserved residues believed to be responsible for domain structure and the interdomain interface^{1–4}. Thus the TCR heterodimer seems similar to that of the Fab portion of immunoglobulins, and its antigen-binding loops, or ‘complementarity determining regions’ (CDRs), have been predicted to lie in a similar overall conformation^{2–4}.

How does the TCR recognize a peptide–MHC complex, and how does this compare to antibody–antigen binding? In the absence of structural data on a TCR–peptide–MHC complex, many possibilities are plausible. These include the binding of a TCR to peptide alone (which has been forced by MHC binding to adopt a unique conformation), binding to MHC determinants changed by peptide contact (the ‘altered self’ hypothesis⁵), or a division of labour in which one V domain of the TCR binds peptide and the other binds a portion of the MHC (ref. 6). In addition, a model has been proposed^{3,4,7} in which the predicted CDR3 loops of both chains (encoded by the V(D)J junctions) bind peptide whereas the germline-encoded CDR loops contact the α helices of the MHC molecule, which flank its peptide-binding groove. One compelling reason for this proposal is the extreme concentration of diversity in the junctional region of TCRs³.

Correlation of TCR usage with any of these models has been only partially successful. For most protein antigens, the T-cell response is dominated by a small number of V_α and/or V_β gene segments^{8–19}. Conserved TCR junctional residues have also been observed in many of these responses, and their alteration changes or abolishes the responses^{10,14,16}. But specific peptide–TCR interactions have never been identified. It is possible, for example, that the conserved CDR3 residues could be contacting a peptide-sensitive portion of the MHC.

We report here a new strategy for defining TCR specificity, designed to detect contact points between heterodimers and antigenic peptides bound to MHC molecules. This method is analogous to the classic genetic technique of second-site suppression, in which mutants in one locus that disrupt

[†] Present address: Clinical Immunology Center, Sydney NSW 2006, Australia.

protein-protein contacts may be compensated for by a change in the second (interacting) locus^{20,21}. It involves immunizing mice with peptides that vary at residues that affect T-cell recognition but not MHC binding, and then analysing both the overall T-cell response to each variant peptide and the TCR sequences of hybridomas that specifically recognize it. By immunizing mice transgenic for either the α or the β chain of a particular heterodimer, we can impose half this TCR on the responding cells. Peptides with changes at one position of the cytochrome peptide evoke better responses in the TCR $_{\beta}$ than in the TCR $_{\alpha}$ mice, but for peptides with substitutions at another site only 3 amino acids away, the reverse is true. This suggests that different

parts of the peptide can be recognized independently by the V $_{\alpha}$ or V $_{\beta}$ chains. Analysis of hybridoma sequences localizes this effect further to CDR3 residues, and indicates that these make direct TCR-peptide contacts. We also see differences in V $_{\beta}$ usage in response to variant peptides, and correlations of CDR3 motifs with specific V $_{\beta}$ chains; these may indicate differences between the framework structures of different V $_{\beta}$ domains. The approach we use here is generally applicable to mapping both TCR-peptide and antibody-antigen interactions. It has also enabled us critically to test some of the models discussed above to obtain a clearer picture of T-cell recognition.

Residues affecting T-cell recognition

The lysine at position 99 of pigeon or moth cytochrome *c* (PCC or MCC; Fig. 1a) is critical for T-cell recognition but does not affect presentation by the I-E^k molecule²². Residues 101 and 102 (leucine and threonine) also affected T-cell recognition but their effects on presentation through I-E^k were equivocal. Finally, initial results suggested that residue 95 (isoleucine) is a T-cell determinant, but subsequent work showed that it has greater effects on interaction with MHC (ref. 23) and possibly on antigen processing²⁴.

We have recently developed a soluble I-E^k peptide-binding assay (P.A.R. *et al.*, manuscript in preparation, ref. 25), which enables us to directly distinguish the effect of an amino-acid

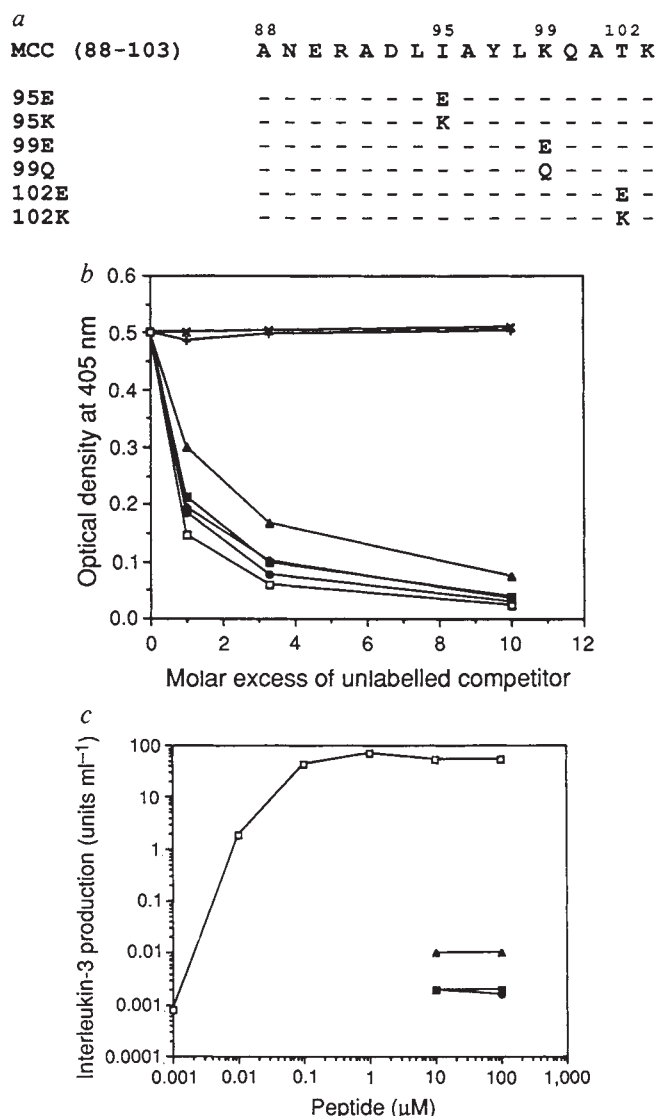


FIG. 1 a, C-terminal peptide from moth cytochrome *c* (MCC) and single amino-acid variants used in this study. Dashes, residues identical to those of MCC 88-103. b, Competition assay for peptide binding to soluble I-E^k. MCC 88-103 ($\square$) and variant peptides 95E (+), 95K ($\times$), 99E ($\blacksquare$), 99Q ($\blacktriangle$), 102E ($\bullet$), and 102K ($\blacklozenge$) were tested for their ability to inhibit signal from an indicator peptide. c, Response of the parental 5C.C7 T-cell line to peptides that bind MHC (symbols as for b). METHODS. a, Peptides were synthesized by standard Fmoc chemistry on a Milligen 5500. b, 50 nM soluble I-E^k was incubated with 1.5 μ M biotinylated MCC, with unlabelled peptide in molar excess as indicated, at pH 5.0 and 37 °C for 2 days. The amount of biotinylated peptide specifically associated with I-E^k was determined by 'capture' ELISA using antiserum specific for the PI-linkage of soluble I-E^k as previously described²⁵. c, 10⁵ 5C.C7 cells, antigen, and 5 $\times$ 10⁴ I-E^k-transfected CHO antigen-presenting cells²⁵ (N1.A4 line) were incubated in 0.2 ml for 20 h. Supernatants were analysed for interleukin-3 content as previously described²⁵.

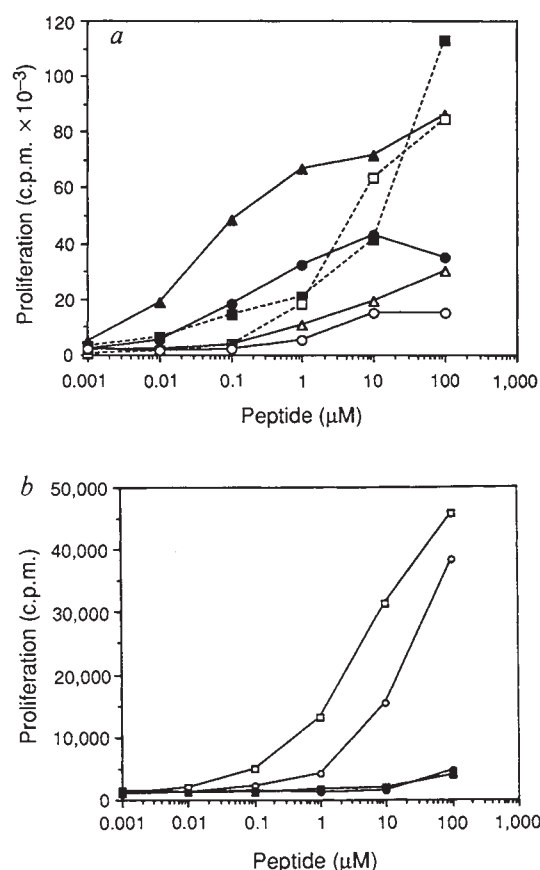


FIG. 2 *In vitro* proliferation of lymph node cells from 5C.C7 TCR α or β chain transgenic mice in response to variant peptides. a, Responses of TCR $_{\alpha}$ (open symbols) or TCR $_{\beta}$ (filled symbols) mice to MCC 88-103 ($\square$, $\blacksquare$, dashed lines), 99Q ($\triangle$, $\blacktriangle$) or 99E ($\circ$, $\bullet$) peptides. b, Responses of TCR $_{\alpha}$ or TCR $_{\beta}$ mice to 102E ($\square$, $\blacksquare$) or 102K ($\circ$, $\bullet$) peptides. METHODS. H-2^{kb} mice were immunized with 100 nmol peptide in complete Freund's adjuvant in the footpads and the base of the tail. 10-12 days later the draining lymph nodes were removed and dissociated. Cells (10⁵) were cultured with antigen in 0.2 ml for 72 h followed by an overnight pulse with 1 μ Ci ³H-thymidine. Wells were collected onto glass-fibre filters and ³H incorporation was determined.

substitution on MHC binding from its effect on T-cell recognition. Accordingly, we synthesized peptide analogues of MCC 88–103, substituted at residues 95, 99 or 102 (Fig. 1a). Each variant was then tested for its ability to compete with a biotinylated MCC peptide for binding to I-E^k. Peptides substituted at 99 and 102 competed effectively for MHC binding, and thus these residues have no obvious effect on the interaction with I-E^k (Fig. 1b). By contrast, both peptides substituted at 95 failed to compete efficiently, and thus these changes must alter or abolish MHC binding. As predicted^{22–24}, residues 99 and 102 are important for T-cell recognition, and the 99 and 102 variants were unable to stimulate the T-cell line 5C.C7 (ref. 26; Fig. 1c).

Disruption of T-cell reactivity

We have established mice transgenic for either the α - or the β -chain TCR genes of 5C.C7, which express their respective transgenes (paired with endogenous β or α chains) on 90% or 98% of their peripheral blood T cells (B.F. *et al.*, manuscript in preparation). These mice were immunized with either the MCC

88–103 peptide or one of the four variant peptides described above which could bind I-E^k (99E, 99Q, 102E or 102K).

As expected, the parental peptide MCC 88–103 elicited strong responses from both 5C.C7 α and β single-chain transgenic mice (Fig. 2a). By contrast, peptides with substitutions at position 99 elicited as good or better responses from the TCR β transgenic mice, and much lower responses from the TCR α mice (Fig. 2a). For peptides substituted at position 102, the situation is reversed, in that the TCR α transgenic mice respond well, whereas the response of the TCR β mice is negligible. Taken together, these data indicate that the 5C.C7 α chain normally interacts with position 99 of MCC 88–103, and that the β chain interacts with position 102.

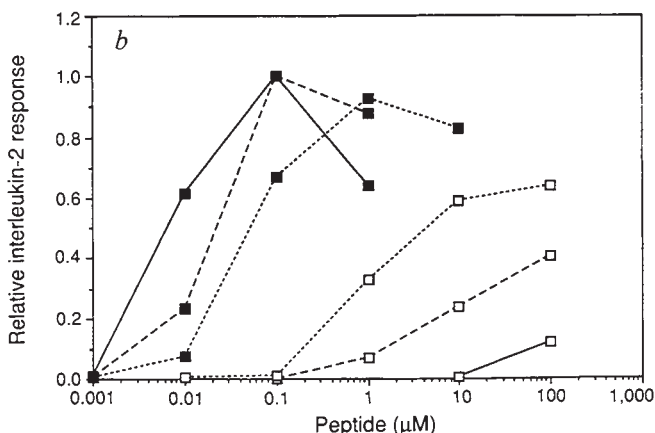
T-cell response to 99E or 99K

T-cell hybridomas were generated from 5C.C7 β chain transgenic animals immunized with either MCC 88–103 or the 99E variant peptide. TCR α chain genes from these hybridomas were analysed by sequencing polymerase chain reaction (PCR)-amplified complementary DNA (Fig. 3). Use of the transgenic

FIG. 3 a, TCR α chain junctional sequences from the parental 5C.C7 line, and from hybridomas from TCR β mice immunized with MCC (88–103) or 99E peptides. All hybridomas shown express V α 11.1 except for hybridoma 104, which expresses V α 11.2. Germline V α 11 sequences are on the left, germline J α sequences are on the right, and N-region nucleotides are centre. The germline J α AF3 sequence has not been determined; the presumptive germline boundary shown is based on shared sequences. Residues at position α 93 are in bold. b, Dose-response curves for hybridomas from a, which recognize both peptides. The ratio of the interleukin-2 response at a given peptide concentration to the maximal response elicited by peptide is shown, for the responses to MCC (88–103) (filled symbols) and 99E (open symbols) of hybridomas 115 (solid line), 226 (dashed line), and 113 (dotted line).

METHODS. a, Mice were immunized and lymph node cells collected as described above. Cells were cultured with 10 μ M peptide and irradiated B10.BR splenocytes (or in later fusions, without added splenocytes) for four days. CD8⁺ cells were killed using complement and supernatant from hybridoma 3-155³⁹. This step significantly increased the fraction of hybridomas which were reactive to peptide, often to more than 90%. Cells were returned to culture in medium supplemented with supernatant from concanavalin A-stimulated rat splenocytes. After 3–4 days, Ficoll-purified cells were fused with BW5147 α β ⁺ (ref. 40) using standard methods⁴¹. Hybridomas were assayed for interleukin-2 production in response to 10 μ M peptide as described²⁵. CD3-bright cells from positive wells were single-cell cloned on a Becton-Dickinson FACStar Plus. Total RNA was prepared by guanidinium isothiocyanate/phenol extraction⁴². Oligo-dT primed cDNA was synthesized using Superscript reverse transcriptase (BRL). PCR amplification was done in Tris/ammonium sulphate buffer⁴³ but without DMSO, using oligonucleotides for V α 11 (GATCTGCGCGCGCTGTCTACTTG-GCTTCAGGAACAAG) and C α (TCAACTGGACACAGCCTCAG). PCR products were run on agarose gels and isolated using GeneClean (Bio 101). Sequencing was done using a nested C α primer (GAACAGGCAGGGTGCTGCTCTGAGA) according to the Sequenase protocol (USB), with the following modifications⁴⁴. Template and primer in Sequenase buffer were annealed by boiling and immediate transfer to dry ice, then warmed to room temperature just before starting extension reactions which were for 2 min at room temperature. Use of the transgenic receptor was confirmed by blocking the responses of representative hybridomas with an anti-V β 3 monoclonal antibody⁴⁵ and by sequencing of PCR-amplified V β 3 cDNA (see Fig. 4). Use of V α 11 was confirmed by blocking with an anti-V α 11 monoclonal antibody⁴⁶. I-E^k restriction was confirmed using I-E^k-transfectant N1.A4 (ref. 25) line as stimulator cells. b, Hybridoma cells (10⁵) were cultured with peptide and 2 $\times$ 10⁴ N1.A4 cells in 0.2 ml for 16–24 h. Supernatants analysed for interleukin-2 as described²⁵.

α					β				
MCC (99K)-reactive TCR									
5C.C7	C A A E				A S N T N K V V F	GTGTRLQVLP			C7
	TGTGCTGCTGAGG				CTTCCAATACCAACAAGTCGTCCTT				
	C A A E				P S N T N K V V F	GTGTRLQVLP			C7
103, 115, 111	TGTGCTGCTGAG	C			CTTCCAATACCAACAAGTCGTCCTT				
	TGTGCTGCTGA	AC			CTTCCAATACCAACAAGTCGTCCTT				
116	C A A E				A S A G N K L T F	GIGTRVLVRP			AC25
	TGTGCTGCTGAG				GCCAGTGCAGGGAACAAGCTAACTTT				
102	C A A D				G N N R I F F	GDGTQLVVKP			MD13
	TGTGCTGCTGA	T			GGCAATAACAGAATCTCTCTT				
99E-reactive TCR									
202, 211, 225, 229	C A A		K		S S G S W Q L I F	GSGTQLTVMP			28
	TGTGCTGCT	AA			ATCTTCTGGCAGCTGGCAACTCATCTTT				
	TGTGCTGC	CAA			ATCTTCTGGCAGCTGGCAACTCATCTTT				
	TGTGCTGC	GAA			ATCTTCTGGCAGCTGGCAACTCATCTTT				
Crossreactive TCR									
113	C A A		G		A S S G Q K L V F	GQGTILKVYL			84
	TGTGCTGC	GGGG			CTTCAAGTGGCCAGAAGCTGGTTTTT				
226	C A A E		C		P S S G Q K L V F	GQGTILKVYL			84
	TGTGCTGCTGAG	C			CTTCAAGTGGCCAGAAGCTGGTTTTT				
104 (V α 11.2)	TGTGCTGCTGAG	C			CTTCAAGTGGCCAGAAGCTGGTTTTT				
114	C A A		V		P A S L G K L Q F	GTGTQVVVTP			AF3
	TGTGCTGCTG	T			ACCTGCCAGTTTGGGGAAGCTCAGTTT				
223	C A A		P		P A S L G K L Q F	GTGTQVVVTP			AF3
	TGTGCTGCT	CC			ACCTGCCAGTTTGGGGAAGCTCAGTTT				



a
MCC (102T)-reactive TCR

		V β		J β
5C.C7	3	C A S S L TGTGCCAGCAGTCTG	N N AACAATG	A N S D Y T F GSGTRLLVI CAAACCTCCGACTACACCTTC 1.2
MCC6	3	C A S S L TGTGCCAGCAGTCT	N N TAACAATG	A N S D Y T F GSGTRLLVI CAAACCTCCGACTACACCTTC 1.2
MCC1	3	C A S S L TGTGCCAGCAGTCTG	N S AACAATG	A N S D Y T F GSGTRLLVI CAAACCTCCGACTACACCTTC 1.2
MCC2	3	C A S S TGTGCCAGCAGT	K N W G AAAAAAGTGGGGG	Q D T Q Y F GPGTRLLVI CAAGACACCCAGTACTTT 2.5

FIG. 4 TCR β -chain junctional sequences from the parental 5C.C7 line, and from hybridomas from α chain mice immunized with *a*, MCC 88–103; *b*, 102E; or *c*, 102K peptides. Germline V and J, and N-region nucleotides are shown as in Fig. 3. Potential D-region nucleotides are underlined. Note that our J β 1.5 sequences have an extra G nucleotide compared with the published sequence⁴⁷, in the alanine codon at β 107/8. The germline V β 6 sequence has not been determined; the presumptive germline boundary shown is based on shared sequences among these hybridomas. Residues that may directly contact peptide are in bold.

METHODS. Hybridoma, RNA and cDNA preparation and PCR and sequencing analysis were essentially as described above. RNA was prepared from original hybridoma cultures. PCR was done using oligonucleotides for C β (GCCGGTTCTAGAAAC-AAGGAGACCTTGGGTGG, used in cloning, or AAGCCTGTGGCCAGGCACAC (a gift from L. Timmerman), used in direct sequencing) and various V β chains (V β 1: GCGTGAATTCGAAACAGCAC-TCATGAACAC; V β 3: GCGTGAATTCGAAGATAT-CTGGTGAAGGGC; V β 6: GCGTGAATTCACATG-GTGATGGTGGC; V β 8.1,2,3: GCGTGAATTC-CTG(T/G)G(T/A)(A/G)CAAAACACATGGAG; V β 14: CTGGAGCTGAATTCGCTCAGACTATCCATCAATGG (a gift from J. Danska and G. Ruberti)). V β 1 and V β 8 sequences were derived by cloning PCR products into Bluescript SK⁺ plasmid (Stratagene) and sequencing double-stranded plasmid DNA using the T7 primer. The other V β sequences were derived by directly sequencing PCR products, as described above. Use of the transgenic receptor chain was confirmed by blocking responses of representative hybridomas with the anti-V α 11 antibody and by sequencing of PCR-amplified V α 11 cDNA. Use of the V β chains identified by PCR was confirmed by blocking responses of representative hybridomas with antibodies against V β 3, V β 4 (ref. 48), V β 6 (ref. 49), V β 8 (ref. 50), and V β 14 (ref. 51). No antibody was available for V β 1. I-E^k-restriction was demonstrated using N1.A4 cells.

b
102E-reactive TCR

		V β		J β
E35	8.1	C A S S E TGTGCCAGCAGTGA	R T G V ACGGACAGGGT	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E3	8.1	C A S S TGTGCCAGCAGTG	G R T G V CTCCGACAGGGT	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E18	8.1	C A S S TGTGCCAGCAGT	K R T G V AAGAGGACAGGGT	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E7	8.1	C A S S D TGTGCCAGCAGTGAT	R T G A CGGACAGGGG	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E22.1	8.1	C A S S D TGTGCCAGCAGTGAT	R A G A CGGGACAGGGAC	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E21	8.2	C A S G E TGTGCCAGCGGTGA	N R AAACCG	S G N T L Y F GEGSRLIVV TTCTGGAATACGCTGTATTTT 1.3
E34	6	C A S S TGTGCCAGCAGTA	I W T G K R D TATGACAGGGGAAAAGAGATT	F S N E R L F F GHGTRLSVL TTTCCACGAAAGATTATTTTTC 1.4
E24	6	C A S S TGTGCCAGCAGTA	T Y T G K G CTTATACAGGGAAGGG	N Q A P L F GEGTRLSVL CAACACGGCTCCGCTTTTTC 1.5
E12	6	C A S S TGTGCCAGCAGTA	R T Q G K G GGACACAGGGGAAGGG	N Q A P L F GEGTRLSVL CAACACGGCTCCGCTTTTTC 1.5
E22.2	6	C A S S TGTGCCAGCAGTA	N H L G R H N ACCATCTCGGACGGCATAA	N Q A P L F GEGTRLSVL CAACACGGCTCCGCTTTTTC 1.5
E26	6	C A S S TGTGCCAGCAGT	S W T G K TCTGACAGGGGAAG	S A E T L Y F GSGTRLTVL AGTGCAGAAACGGCTGTATTTT 2.3
E13	6	C A S S TGTGCCAGCAGTA	I G R L P TAGGCAGACTCCCT	S Q N T L Y F GSGTRLTVL AGTCAAAACACCTTGTACTTT 2.4
E20	1	C A S S Q TGTGCCAGCAGCCAA	G Q G R G GGACAGGGGAGGGGG	T E V F F GKGTRLTVV ACAGAAAGTCTTCTTTT 1.1
E28	1	C A S S Q TGTGCCAGCAGCCAA	G Q G V G GGACAGGGGAGTAGG	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E30	1	C A S S TGTGCCAGCAGCC	P G Q G T GGGACAGGGGAC	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
E6	1	C A S S TGTGCCAGCAGC	N AAC	N S D Y T F GSGTRLLVI AACTCCGACTACACCTTC 1.2
E4 (K)	1	C A S TGTGCCAGC	G Q G GGACAGGGGA	N N S P L Y F AAGTRLTVT ATATTCCGCCCTGTACTTTT 1.6
E8	4	C A S S TGTGCCAGCAGC	G T GGGACAA	T N S D Y T F GSGTRLLVI CAAACCTCCGACTACACCTTC 1.2

c
102K-reactive TCR

		V β		J β
K10	14	C A W S TGTGCCTGGAGTC	P G T G CCGGACAGGGC	H T E V F F GKGTRLTVV ACACAGAAGTCTTCTTTT 1.1
K5	14	C A W S TGTGCCTGGAGTC	P P S G CCCCCTCAGGGC	Q T E V F F GKGTRLTVV ACACAGAAGTCTTCTTTT 1.1
K17	14	C A W S TGTGCCTGGAGTC	P P E G CCCCCTGAAGGGC	Q T E V F F GKGTRLTVV ACACAGAAGTCTTCTTTT 1.1
K4	14	C A W S L TGTGCCTGGAGTCT	E G G R A AGAGGGGGGGGGGGC	S D Y T F GSGTRLLVI CTCCGACTACACCTTC 1.2
K7	14	C A TGTGCC	G T V R D Q GGACCGCTCCGGACACAA	Y A E Q F F GPGTRLTVL TATGCTGAGCAGTTCTTTC 2.1
K20	14	C A W S TGTGCCTGGAGTC	Q G R E ACGGACAGGAGG	D T L Y F GAGTRLSVL ACACCTTGTACTTTT 2.4
K14	14	C A W S TGTGCCTGGAGT	F G A TTTGGAGC	Q D T Q Y F GPGTRLLVI CCAAGACACCCAGTACTTT 2.5
K6	8.3	C A S S TGTGCCAGCAGTG	A F W G GTTTCTAGGGT	Y A E Q F F GPGTRLTVL TATGCTGAGCAGTTCTTTC 2.1
K19	6	C A S S TGTGCCAGCAG	P N W G G N TCCGACCTGGGGGAGGAAT	S A E T L Y F GSGTRLTVL AGTGCAGAAACGGCTGTATTTT 2.3

β chain, and of α chains identified by PCR, was confirmed by blocking antigen responses with anti- $V_{\beta 3}$ or anti- $V_{\alpha 11}$ antibodies and by PCR as described in the legend to Fig. 3.

Interaction of the 5C.C7 α chain with position 99 of the peptide, suggested by the data in Fig. 2b, is even more strongly indicated by these TCR sequences. The receptors on MCC-reactive hybridomas are similar to those previously described⁹, using $V_{\alpha 11.1}$ and preserving the germline $V_{\alpha 11}$ -encoded glutamate at $\alpha 93$, at the V-J junction, or in one case conservatively substituting an aspartate at this position. The 99E-specific TCR also use the $V_{\alpha 11.1}$ gene segment, and thus differ from the MCC-reactive TCR only in their V-J junctions and J_{α} segments. All eight 99E-specific hybridomas analysed have the same amino-acid sequence whereas at the DNA level at least three independent clones are evident. Strikingly, the germline-encoded, negatively charged residue at $\alpha 93$ evident in the MCC-reactive receptors has been replaced with a positively charged lysine (Fig. 3a). The germline sequence of the $J_{\alpha 28}$ segment (B. Koop and L. Hood, personal communication) shows that this codon is created largely by random nucleotide (N-region) addition, suggesting a strong selection for lysine at this position. Seven additional 99E-specific hybridomas were screened by PCR with a $J_{\alpha 28}$ oligonucleotide terminating in a lysine codon and were all positive, indicating that this is the predominant sequence.

Thus a lysine to glutamate change at position 99 of the peptide seems to select for a reciprocal charge reversal at $\alpha 93$ of the TCR, suggesting a close interaction of these residues, as in a 'salt-bridge' (or charge-charge pair). Specific recognition of 99E seems to require a different J_{α} segment, which may position residue $\alpha 93$ for optimal peptide contact, or may itself contact the peptide as well.

A few hybridomas derived from both immunizations responded to either MCC 88-103 or the 99E peptide. The relative responses to MCC versus 99E of these hybridomas show the importance of both V_{α} - J_{α} junctional residues and J_{α} segments in determining specificity. Hybridomas 113 and 226 differ only in two residues at the V-J junction (Fig. 3a); hybridoma 113, with a glycine-alanine sequence at $\alpha 93$ -94 responds much better to 99E versus MCC 88-103 than does hybridoma 226, with a glutamate-proline sequence (Fig. 3b). Hybridoma 226 nonetheless responds much better to 99E than does hybridoma 115, which differs only in its J_{α} .

Response to substitutions at 102

Hybridomas were made from 5C.C7 α transgenic animals immunized with either the 102E or 102K peptides, or with MCC 88-103 (102T). The first two peptides produced a high frequency (about 50%) of crossreactive hybridomas that responded to both peptides at 10 μ M. Analysis focused on the remaining hybridomas that specifically responded to the immunizing peptide at 10 μ M.

In these experiments, six of seven hybridomas that responded to MCC 88-103 express $V_{\beta 3}$ (see Table 1). By contrast, none of the 102E- or 102K-reactive hybridomas use $V_{\beta 3}$, but rather each set expresses a total of three to five different V_{β} segments. Three of three MCC-reactive hybridomas analysed have TCRs with an asparagine at residue $\beta 100$ at the V-D junction, encoded partially by N-region nucleotides (Fig. 4a), fitting the pattern seen previously by Hedrick *et al.*⁹

In the sequences of 102E-reactive hybridomas, we again see evidence for direct CDR3-peptide contacts (Fig. 4b). In particular, the $V_{\beta 8.1}$ -expressing hybridomas all have an arginine at $\beta 100$, at the V-D junction. The codons for these arginines are N-region derived, suggesting that these residues have been highly selected for recognition. Hybridoma 102E21, which expresses $V_{\beta 8.2}$ and a different J_{β} segment, also has an N-region encoded junctional arginine, at $\beta 101$. All the $V_{\beta 6}$ -expressing hybridomas similarly have at least one N-region encoded lysine or arginine in their V_{β} -D β - J_{β} junctions. Thus, in many

examples, introducing the negatively charged glutamate into the cytochrome peptide at 102 gives rise to a conserved, N-region encoded, positively charged residue at the β -chain V-D-J junction. This again suggests a direct interaction with the peptide through salt-bridges.

A very different CDR3 sequence motif is exhibited by the 102E-reactive hybridomas which express $V_{\beta 1}$. These show a conserved glutamine at $\beta 101$ (or the closely related asparagine in one case). Either of these could interact with the glutamate at 102 through a strong hydrogen bond²⁷. In fact, charged residues in known antibody-antigen interfaces are more often involved in hydrogen bonds than in salt-bridges²⁸. The threonines in the junctions of the $V_{\beta 4}$ sequence may be involved in similar hydrogen bonds.

Most hybridomas reactive to the second 102 peptide variant, 102K, express $V_{\beta 14}$ (Table 1). Several of these have negatively charged residues at $\beta 100$, 101 or 102 in their V-D-J junctions (Fig. 4c), which could be involved in charge-charge pairs with the lysine at 102 of the peptide. By contrast, none of the 18 102E-reactive TCR have negatively charged residues at these positions. Three 102K-reactive hybridomas (K6, K14 and K19), expressing three different V_{β} segments, have one or more aromatic residues in their V-D-J junctions, whereas relatively few such residues are found in the junctions of the 102E-reactive hybridomas. Aromatic side chains are thought to have ring faces with a net negative charge, which may be able to interact with the positively charged termini of lysine and arginine side chains^{29,30}. The junctional aromatic residues in the 102K-reactive TCRs thus may have been selected for their contribution to binding peptide.

It is particularly significant that hybridoma K19 shares both $V_{\beta 6}$ and $J_{\beta 2.3}$ (as well as the transgenic 5C.C7 α chain) with a 102E-reactive hybridoma, E26. Neither of these hybridomas crossreacts with the other peptide. Thus, as we found for the TCR α chains of the 99E- or K-reactive hybridomas, changing only the CDR3 loop of the β chain can completely change specificity for peptide.

TABLE 1 V_{β} usage in hybridomas from α transgenic mice

Immunizing peptide	Total independent hybridomas	V_{β}	Number positive
MCC 88-103 (102T)	4-7	3	6
		ND	1
	18	1	4
		3	0
		4	1
		6	6
		8.1	5
		8.2	1
		14	0
		ND	1
102K	9	1	0
		3	0
		4	0
		6	1
		8.3	1
		14	7

V_{β} chain usage by 5C.C7 α -chain hybridomas directed against various peptides. ND, Not determined; these hybridomas were negative by PCR for $V_{\beta 1}$, 3, 4, 6, 8 and 14 expression. Hybridomas were primarily from single animals for 102E and 102K, and from pooled lymphocytes from two animals for MCC 88-103. 102E- and K-reactive hybridomas shown did not react to 10 μ M of 102K or 102E, respectively. Three MCC 88-103-reactive hybridomas were screened by blocking with anti- $V_{\beta 3}$ monoclonals; all other hybridomas indicated as positive had in-frame V_{β} sequences.

Peptide dose-response curves of representative hybridomas from each immunization roughly reflect the responses of the bulk lymph node cells (as shown in Fig. 2). Most hybridomas from TCR β transgenic mice, reactive to either MCC or the 99E peptides, show similar dose-response curves to the parental 5C.C7 line. MCC-reactive hybridomas from TCR α transgenic mice also show similar, high apparent avidities. Hybridomas from these mice reactive to the 102E and 102K peptides show generally lower apparent avidities, with dose-response curves falling over a wide range.

Topology for TCR-peptide-MHC interaction

Our data strongly indicate that the junctional region of 5C.C7 V α interacts with position 99 of the cytochrome peptide, whereas the junctional region of V β interacts with position 102. Assuming an antibody-like structure for the $\alpha\beta$ heterodimer, this allows us to orient the TCR over the peptide as shown in Fig. 5. In addition, the data of Ronchese *et al.*³¹ showed that a change at residue 29 of the E β molecule, near one end of the putative binding pocket of I-E k , prevented the presentation of some cytochrome *c* peptides that differ at their C-termini. This suggests that the N-terminus of the cytochrome peptide extends to the left (in Fig. 5) to make maximal use of the peptide-binding site. Recent data of our own with MCC/I-E k -specific antibodies also supports this proposed orientation (P.A.R. *et al.*, manuscript in preparation) and it is consistent with the proposal of Sette coworkers regarding other E k -restricted peptides³². The resultant ternary complex would therefore position the germline-encoded portion of V α over the α helix of E β , and V β over the α helix of E α .

Discussion

The reciprocal charges that we elicit in the predicted CDR3 regions of both the α and the β chains of TCRs that respond to variant peptides constitute the strongest evidence of direct TCR-peptide contact. The lack of reciprocal charges in every case should not be surprising, as charged residues in antigen-antibody interfaces are often instead involved in hydrogen bonds, as noted above. The TCR-peptide contacts we identify are relatively independent, as changes in the peptide can often be accommodated solely by changes in CDR3 loops of either the α or the β chain. These data support models of T-cell

recognition in which these regions play the major part in peptide recognition^{3,4,7}, and suggest an orientation of the TCR over the antigen-MHC complex (Fig. 5). This topology provides us with a working model of cytochrome *c* recognition, and may be generalized if TCR can only interact with MHC molecules in one orientation (although specific TCR-MHC contacts vary greatly, as discussed below).

These data also indicate that V β gene segments are not interchangeable. Responses to MCC peptides with different residues at position 102 have very different, often mutually exclusive, patterns of V β usage (Table 1). CDR sequence motifs are closely linked to the particular V β segments used, even when the TCR α chain, peptide and MHC are the same. Germline V β segments differ in the presumed MHC contact loops, as well as in residues predicted to lie at the interface between the α and β chains, which may affect the overall configuration of the heterodimer³³. These differences may result in shifts of the CDR3 loops along or across the peptide in the MHC groove, or in proximity to a side chain on the peptide. This last possibility may be of particular significance here, as salt bridges involve the charged ends of the longest side chains and thus may require more distance between the TCR and the peptide to be useful. This may account for the change in V β usage when the threonine at 102 (proposed to bond with asparagine at β 100) is replaced with the longer glutamate. By contrast, when the lysine at 99 is replaced with glutamate we see no change in V α usage, perhaps because the overall length of the interacting pair of residues need not change.

Different TCRs can interact differently with the same MHC molecule as indicated by the fact that different T cells are blocked by different subsets of mutations in the MHC, at residues predicted to point toward the TCR (refs 34 and 35, and E. Ehrlich *et al.*, manuscript in preparation). Some positive or negative interaction between the peptide and germline V α and V β residues may occur, but it is difficult to envision a major role for these sequences in peptide contact in most of our examples, as this would not account for the reciprocal charge motifs we see in CDR3 sequences. In addition, particularly in the case of V β , there are very few (~10) germline sequences in some strains of wild mice³⁶ and in domestic chickens³⁷, which must be used to recognize a very large number of possible peptides.

These indications of variations between V β structures help to explain the failure of CDR3 transfers between different V β s to confer specific peptide reactivities (P. A. Patten, E. A. Rock, T. Sonoda, B. F. de StG. and M.M.D., manuscript submitted). They also may bear on the restricted V segment usage seen in the response to many peptides⁹⁻¹⁹. Thus, in some cases only one V α :V β combination may position the CDR3 loops properly for recognition of a particular peptide bound to MHC.

In our experiments, changing a single residue in the antigenic peptide changes the degree of heterogeneity of the V β response to a surprising degree (Table 1). Glutamate and lysine have very flexible side chains compared with the original threonine at 102, and may be able to interact with CDR3 loops positioned differently by different V β chains. Thus the diverse T-cell responses observed against some peptides (ref. 38, and J. A. Kobari, L. Hood and N. Shastri, manuscript in preparation) may be because of greater flexibility in their exposed side chains. *Note added in proof:* We have recently obtained a hybridoma from a TCR α transgenic mouse which expresses V β 8.1 and J β 2.3, and specifically recognizes the 102K peptide. This provides a second example, with the V β 8.1-expressing 102E-reactive hybridomas, of a pair of TCRs which differ only in the CDR3 loops and J segments of the TCR β chain and have non-overlapping specificities. □

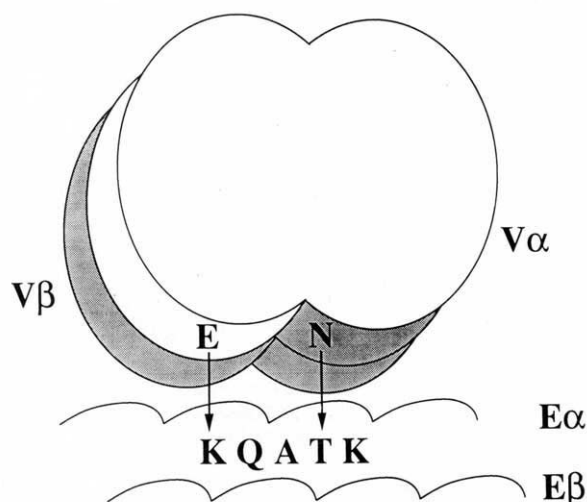


FIG. 5 Proposed configuration of the 5C.C7 TCR/MCC 88-103/I-E k complex. Shown is a schematic side view of the interface, with the T cell above and the antigen-presenting cell below. MCC residues 99-103 are shown in the presumptive peptide-binding groove, flanked by the α helices of I-E k . The V β domain of the TCR is shaded. The CDR1 and 2 loops of both domains are shown over the α helices of the I-E k heterodimer (E k α E k β) contacting I-E k , and the proposed peptide contact residues in the CDR3 loops are shown in bold.

Received 5 July; accepted 23 October 1991.

1. Kronenberg, M., Siu, G., Hood, L. E. & Shastri, N. A. *Rev. Immun.* **4**, 529-591 (1986).
2. Novotny, J., Tonegawa, S., Saito, H., Kranz, D. M. & Eisen, H. N. *Proc. natn. Acad. Sci. U.S.A.* **83**, 742-746 (1986).
3. Davis, M. M. & Bjorkman, P. J. *Nature* **334**, 395-402 (1988).
4. Chothia, C., Boswell, D. R. & Lesk, A. M. *EMBO J.* **7**, 3745-3755 (1988).

5. Zinkernagel, R. M. & Doherty, P. C. *Nature* **251**, 547–548 (1974).
6. Saito, T. & Germain, R. N. *Nature* **329**, 256–259 (1987).
7. Claverie, J. M., Prochnicka-Chalouf, A. & Bougueleret, L. *Immun. Today* **10**, 10–14 (1989).
8. Hedrick, S. M. *Adv. Immun.* **43**, 193–234 (1988).
9. Hedrick, S. M. *et al. Science* **239**, 1541–1544 (1988).
10. Engel, I. & Hedrick, S. M. *Cell* **54**, 473–484 (1988).
11. Acha-Orbea, H. *et al. Cell* **54**, 263–273 (1988).
12. Urban, J. L. *et al. Cell* **54**, 577–592 (1988).
13. Rodewald, H. R., Koszinowski, U. H., Eichmann, K. & Melchers, I. *J. Immun.* **143**, 4238–4243 (1989).
14. Lai, M.-Z., Jang, Y.-J., Chen, L.-K. & Geffer, M. L. *J. Immun.* **144**, 4851–4856 (1990).
15. Sorger, S. B., Paterson, Y., Fink, P. J. & Hedrick, S. M. *J. Immun.* **144**, 1127–1135 (1990).
16. Danska, J. S., Livingstone, A. M., Paragas, V., Ishihara, T. & Fathman, C. G. *J. Exp. Med.* **172**, 27–33 (1990).
17. Aebischer, T., Oehen, S. & Hengartner, H. *Eur. J. Immun.* **20**, 523–531 (1990).
18. Yanagi, Y., Maekawa, R., Cook, T., Kanagawa, O. & Oldstone, M. B. A. *J. Virol.* **64**, 5919–5926 (1990).
19. Wither, J., Pawling, J., Phillips, L., Delovitch, T. & Hozumi, N. *J. Immun.* **146**, 3513–3522 (1991).
20. Jarvik, J. & Botstein, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2738–2742 (1975).
21. Botstein, D. & Maurer, R. A. *Rev. Genet.* **16**, 61–83 (1982).
22. Fox, B. S. *et al. J. Immun.* **139**, 1578–1588 (1987).
23. Schwartz, R. H. *et al. J. Immun.* **135**, 2598–2609 (1985).
24. Fox, B. S., Carbone, F. R., Germain, R. N., Paterson, Y. & Schwartz, R. H. *Nature* **331**, 538–540 (1988).
25. Wettstein, D. A., Boniface, J. J., Reay, P. A., Schild, H. & Davis, M. M. *J. exp. Med.* **174**, 219–228 (1991).
26. Fink, P. J. *et al. Nature* **321**, 219–226 (1986).
27. Fersht, A. R. *et al. Nature* **314**, 235–238 (1985).
28. Davies, D. R., Padian, E. A. & Sheriff, S. A. *Rev. Biochem.* **59**, 439–473 (1990).
29. Burley, S. K. & Petsko, G. A. *FEBS Lett.* **203**, 139–143 (1986).
30. Dougherty, D. A. & Stauffer, D. A. *Science* **250**, 1558–1560 (1990).
31. Ronchese, F., Schwartz, R. H. & Germain, R. N. *Nature* **329**, 254–256 (1987).
32. Krieger, J. I. *et al. J. Immun.* **146**, 2331–2340 (1991).
33. Colman, P. M. *Adv. Immun.* **43**, 99–132 (1988).
34. Ajitkumar, P. *et al. Cell* **54**, 47–56 (1988).
35. Peccoud, J., Dellabona, P., Allen, P., Benoist, C. & Mathis, D. *EMBO J.* **9**, 4215–4223 (1990).
36. Huppi, K. E. *et al. Immunogenetics* **27**, 51–56 (1988).
37. Tjoelker, L. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 7856–7860 (1990).
38. Taylor, A. H., Haberman, A. M., Gerhard, W. & Caton, A. J. *J. exp. Med.* **172**, 1643–1651 (1990).
39. Sarmiento, M., Glasebrook, A. L. & Fitch, F. W. *J. Immun.* **125**, 2665–2672 (1980).
40. White, J. *et al. J. Immun.* **143**, 1822–1825 (1989).
41. Rock, K. in *Hybridoma Technology in the Biosciences and Medicine* (ed. Springer, T. A.) 527–544 (Plenum, New York, 1985).
42. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
43. Chen, S. & Evans, G. A. *Biotechniques* **8**, 32–33 (1990).
44. Casanova, J.-L., Pannetier, C., Jaulin, C. & Kourilsky, P. *Nucleic Acids Res.* **18**, 4028 (1990).
45. Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335**, 796–801 (1988).
46. Jameson, S. C. *et al. J. Immun.* **147**, 3185–3193 (1991).
47. Gascoigne, N. R. J., Chien, Y., Becker, D. M., Kavalier, J. & Davis, M. M. *Nature* **310**, 387–391 (1984).
48. Tomonari, K., Lovering, E. & Spencer, S. *Immunogenetics* **31**, 5–6 (1990).
49. Kanagawa, O., Palmer, E. & Bill, J. *Cell. Immun.* **119**, 412–426 (1989).
50. Staerz, U., Rammensee, H., Benedetto, J. & Bevan, M. J. *Immun.* **134**, 3994–4000 (1985).
51. Liao, N.-S., Maltzman, J. & Raulet, D. H. *J. exp. Med.* **170**, 135–143 (1989).

ACKNOWLEDGEMENTS. We thank R. Schwartz, P. Bjorkman, J. Danska and G. Ruberti for discussions, B. Koop and L. Hood for the communication of the J_α 28 sequence before publication, O. Kanagawa, D. Raulet and K. Tomonari for gifts of antibodies, W. Born for the use of the hybridoma fusion partner and B. Robertson for preparation of the manuscript. This work was supported by the NIH. J.L.J. is supported by the Medical Scientist Training Program and P.A.R. by a fellowship from Merck Sharp and Dohme.

LETTERS TO NATURE

Interpretation of solar-cycle variability in high-degree p-mode frequencies

David J. Evans* & B. Roberts†

* Astronomy Unit, School of Mathematical Sciences, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK
 † Department of Mathematical and Computational Sciences, University of St Andrews, St Andrews, Fife KY16 9SS, UK

RECENTLY Libbrecht and Woodard¹ and Elsworth *et al.*² have demonstrated that the frequencies of solar acoustic p-mode oscillations vary significantly over the solar cycle. We have previously suggested that cyclic variations in the magnetic activity of the Sun could modulate the p-mode frequencies in a similar way. In particular, we investigated^{3–5} simple models of the ‘magnetic canopy’, which permeates the solar atmosphere and overlies all of the Sun’s surface, to determine its influence on p-mode frequencies. Here we make a comparison of our model predictions with the observations of Libbrecht and Woodard. We find that, despite the simplicity of our model, we are able to obtain good agreement with the observed frequency shifts for modes of frequency less than 4 mHz, through a mechanism in which an increasing magnetic field induces ‘stiffening’ of the Sun’s chromosphere. Above this frequency there is clearly something missing from our model. We speculate that the behaviour above 4 mHz is related to a cutoff frequency in the solar atmosphere, above which waves are trapped only partially.

Helioseismology studies the resonant frequencies of solar acoustic waves (p-modes) trapped within the Sun. The upper turning point of solar p-modes is near the surface and rises with increasing frequency. The depth of the cavity mainly depends on ν/l where ν is the frequency and l is the spherical harmonic degree of the mode. The degree l is related to horizontal wavenumber k by $k = \sqrt{l(l+1)}/R_{\text{Sun}}$, R_{Sun} being the solar radius (696 Mm). Low- l modes ($l < 10$) are the most deeply penetrating, whereas high- l modes ($l > 100$) are confined to the layers near the surface. Modes trapped near the surface are most likely to be affected significantly by changes in the solar atmosphere. High- l modes have small horizontal wavelengths relative to the solar radius. As we are primarily interested in effects due to the solar atmosphere, it is reasonable to adopt a plane-parallel

model in which the curvature of the Sun is ignored. We have considered^{3–5} a simple, field-free polytropic model surmounted by an isothermal chromosphere (the region between the surface and the hot corona) permeated by a horizontal magnetic field. We assume that the chromosphere begins at the temperature minimum (~500 km above the visible surface) and that waves propagate parallel to the magnetic field.

For two simple cases of the chromospheric magnetic field we have been able to derive an asymptotic dispersion relation valid for small horizontal wavenumber k . Here we discuss the case of a uniform magnetic field⁴ because this produces an increase in frequency with increasing magnetic activity, in accordance with the observations reported in ref. 1. Then we find that the cyclic frequency ν of a p-mode of order n is given, approximately, by⁴

$$\nu \approx \frac{\omega_*}{2\pi} \left(1 + \psi(n, k) \frac{B_c^2}{2\mu P_p} (2kz_0)^{m+1} \right) \quad (1)$$

where ω_* , the angular frequency in the absence of chromospheric effects, magnetic or otherwise, is determined from

$$\frac{m+1}{\gamma} \frac{\omega_*^4}{g^2 k^2} - (m+2n) \frac{\omega_*^2}{gk} + \left(m - \frac{m+1}{\gamma} \right) = 0 \quad (2)$$

and

$$\psi = \frac{\Gamma(1+m+n) \left(\frac{\omega_*^2}{gk} + 2 - \frac{gk}{\omega_*^2} \right)}{\left(\frac{m+1}{\gamma} \left[1 + \frac{g^2 k^2}{\omega_*^4} \right] - m \frac{g^2 k^2}{\omega_*^4} \right) \times \Gamma(m+1) \Gamma(m+2) (n-1)! \left(\frac{\omega_*^4}{g^2 k^2} - 1 \right)} \quad (3)$$

Here γ is the adiabatic index and m is the polytropic index of a gas with density proportional to $(1+z/z_0)^m$, z being the depth below the temperature minimum. The term z_0 is the temperature scale-height in the field-free lower region, where the temperature is $T_p(1+z/z_0)$, T_p being the temperature at the top of the polytrope. Also, g is the acceleration due to gravity, p_p is the gas pressure at the top of the polytrope, B_c is the magnetic field strength in the chromosphere, and μ is the magnetic permeability.

The asymptotic dispersion relation (1) is valid for $l \leq 150$, although it tends to underestimate the frequency shift for high-order modes. It can be seen from equation (1) that, assuming reasonable values for m and γ , the frequency shift for p-modes

is always positive. For given polytropic and adiabatic indices, the frequency ω_* depends solely on the order n and wavenumber k (degree l) of the mode. In fact, for an atmosphere that is stable to convection ($m = 1/[\gamma - 1]$), $\omega_*^2 = (n + \alpha)2gk/m$ where α is the phase change due to the reflection at the upper boundary^{3,6}. For a stably stratified atmosphere, constant magnetic field and $m = 2$, the frequency shift is roughly proportional to $n^{5/2}k^3$ in the limit of large n . The dependence of the frequency shift on the magnetic field is due to the equivalent pressure exerted, $B_c^2/2\mu$. This magnetic pressure modifies the propagation of waves through the fluid. It is as if the fluid is stiffened by the addition of elastic threads. This has the effect of making the upper reflecting boundary of the p-mode cavity more like a rigid wall. Consequently, there is an increase in the value of α and hence in the frequency.

The data of ref. 1 consist of two sets of oscillation frequencies for the same set of pairs (n, l) but derived from observations at two distinct epochs, 1986 and 1988. The year 1986 coincided with solar minimum, whereas 1988 saw increased magnetic activity on the rise to the 1989 solar maximum. We suppose that the increase in magnetic activity manifests itself as a rise in mean chromospheric magnetic field strength from low values (~ 5 G) at solar minimum (1986) to higher values (~ 25 – 30 G) in 1988 (ref. 7).

To compare the predictions of equation (1) with the observations we adopt the following simple procedure. For each mode (that is, each distinct pair n, l) we take as a 'base' frequency the 1986 value and then calculate the change in cyclic frequency $\Delta\nu = (\nu_{88} - \nu_{86})$ between 1986 and 1988. We divide the modes according to their base frequency into bins of width $100 \mu\text{Hz}$ and average the values of $\Delta\nu$ in each bin. We repeat the calculation for the theoretical model by substituting the field-free frequency ν_* ($=\omega_*/2\pi$) for ν_{86} and the frequency as calculated using equation (1) for ν_{88} . We must first choose a suitable set of parameters describing our model polytrope and atmosphere. Here we are constrained by the simplicity of our model. We use averages of the parameters calculated from a theoretical model⁸ of the solar convection zone. The values used are as follows: $l = 1$ – 140 ; $n = 1$ – 27 ; $R_{\text{gas}} = 13087.3 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$ (we assume an ideal gas satisfying $p = R_{\text{gas}}\rho T$, where ρ is density); $\gamma = 1.5$; $m = 1.9$; $p_p = 80.0 \text{ kg m}^{-1} \text{ s}^{-2}$; $T_p = 4,170 \text{ K}$; $z_0 = 577.6 \text{ km}$.

Figure 1 shows a comparison of theoretical and observed frequency shifts. We have plotted curves for three chromospheric field strengths: 5 G (solar minimum), 25 G and 30 G (significant magnetic activity). It is clear that the theoretical effect of a 5-G

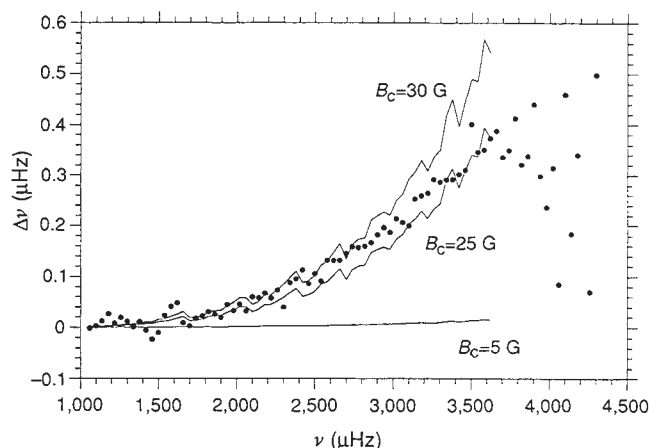


FIG. 1 A comparison of the observed¹ (dots) and theoretical frequency shifts (lines) according to equation (1). For the observations of Libbrecht and Woodard¹, modes were binned according to their 1986 frequency (bin widths of $100 \mu\text{Hz}$) and then the frequency shifts $\Delta\nu = (\nu_{88} - \nu_{86})$ were averaged for each bin to arrive at the final point plotted. For the theoretical results we replace ν_{86} by the frequency predicted for zero magnetic field and ν_{88} by the frequency calculated from equation (1) with a chromospheric magnetic field of strength $B_c = 5$ G, 25 G and 30 G.

field is negligible, so that assuming the solar minimum frequency to be given by the field-free case is reasonable. Below base frequencies of 4 mHz, the agreement between theory and observation is good, despite the simplicity of our model. Above 4 mHz, the observed frequency shifts indicate a rapid drop to zero. Equation (1) does not show such a drop, thus indicating the existence of physical processes in the Sun that become important at high frequencies but are not represented by the simple model we have put forward. This is likely to be due to poor modelling of the upper chromosphere, as only the high-frequency modes are affected by the upper reaches of the solar atmosphere. Our model of uniform magnetic field strength gives an Alfvén speed (the characteristic speed for plasma waves supported by the magnetic field) which increases exponentially with height, clearly unrealistic high above the temperature minimum.

Another interesting result to emerge from the data of ref. 1 is seen after adjusting the frequency shift $\Delta\nu$ to the equivalent value for a base frequency of 3 mHz. Libbrecht and Woodward¹ find that $\Delta\nu$ is proportional to l . We can mimic this procedure in equation (1) by forcing n to be a function of l so that ω_* , as given by equation (2), is a constant. We have done this for the simple case of a solar model that is stable to convection. After some straightforward analysis assuming n to be large ($n > 10$), so that we may approximate the gamma functions in equation (1) using Stirling's formula, we find that

$$\Delta\nu = \frac{1}{2\pi} \frac{1}{\Gamma(m+1)^2} \left(\frac{\omega_*}{\nu_0} \right)^{2m+1} \frac{B_c^2}{2\mu p_p} \frac{H_0 \nu_0}{R_{\text{sun}}} l \quad (4)$$

where $H_0 = c_s^2/\gamma g$, the density scale-height at $z=0$, and $\nu_0^2 = g/[m(m+1)H_0]$. Thus our model gives qualitative agreement

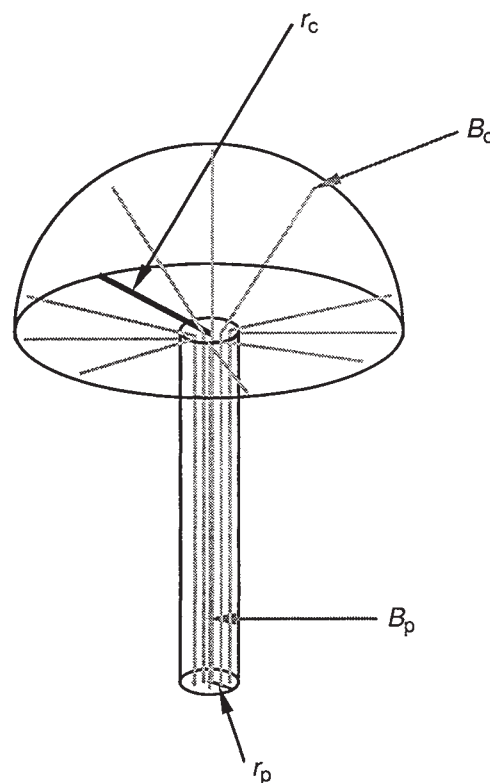


FIG. 2 Slender flux tubes in the photosphere fan out rapidly into the chromosphere, forming a magnetic canopy. Here we have idealized this to a thin cylinder of magnetic flux of radius r_p joined to a hemisphere of radius r_c , the flat face of which represents the base of the chromosphere. The magnetic field in the cylinder is a uniform B_p and over the curved surface of the hemisphere the magnetic field is also uniform. Conservation of flux implies that the total flux in the cylinder is the same as that spread over the curved surface of the hemisphere. Then the mean-square chromospheric field (B_c^2) is simply the average of B_c^2 over the flat surface of the hemisphere.

with the results of ref. 1 in reproducing the observed linear relationship between $\Delta\nu$ and l at fixed frequency. This behaviour is to be expected for perturbations confined to the surface layers. (A variational approach⁹ to this problem indicates that surface-layer perturbations cause frequency shifts that are inversely proportional to the mode inertia $\sim \int \rho \xi \cdot \xi \, dV$, where ξ is the displacement eigenfunction and ρ is the gas density. For modes with constant frequency and confined to the outer parts of the Sun, mode inertia is proportional to $1/l$ (ref. 1)). It should be noted that the observational plot of $\Delta\nu$ against l (Fig. 2b of ref. 1) shows a nonzero intercept on the $\Delta\nu$ axis when $l=0$, whereas here our calculated $\Delta\nu$ is zero for $l=0$. This difference is a consequence of geometry: the mode inertia for a plane-parallel polytrope is unbounded as l tends to zero, but in a sphere it remains finite.

Goldreich *et al.*¹⁰ have also discussed p-mode frequency shifts in terms of a magnetic model. They conclude that the frequency increases at solar maximum are due to the increased density of intense vertical magnetic flux tubes just below the photosphere (the visible surface). This again will tend to 'stiffen' the fluid, thus increasing propagation speeds. Gough and Thompson¹¹ looked at similar ideas in trying to explain frequency splitting data, and Dziembowski and Goode¹² have used inversion techniques in an attempt to verify such a model. Goldreich *et al.* find that it is possible to reproduce the observed frequency shifts provided that the flux tubes contain magnetic fields that remain strong to a greater depth than would otherwise be expected, assuming, as is currently believed, that flux tubes are formed by convective collapse¹³.

Woodard *et al.*¹⁴ have shown that observed frequency shifts, for both long- and short-term changes in solar activity, are proportional to the mean absolute photospheric magnetic field as measured by magnetograms. Does the chromospheric model presented here also have this property? It can be seen from equation (1) that the frequency shift depends on the square of the chromospheric magnetic field strength. Over a large area, in which the chromospheric magnetic field varies, the important parameter in our model is the mean square magnetic field at the base of the chromosphere, and we must relate this to the mean absolute photospheric field.

Consider a simple model of an isolated magnetic flux tube expanding outwards from the chromosphere^{13,15} (see Fig. 2). The photospheric part of the tube is taken to be a slender cylinder of radius r_p and uniform field strength B_p . We assume that in the chromosphere this field fans out radially and uniformly over the surface of a hemisphere of radius r_c . We also assume that all photospheric flux appears in the form of identical slender flux tubes and that variations in chromospheric magnetic field strength are due entirely to changes in the number of flux tubes per unit area. Then it is simple to show that the mean square chromospheric magnetic field, $\langle B_c^2 \rangle$, is related to the mean absolute photospheric magnetic field, $\langle B_{ph} \rangle$, by

$$\langle B_c^2 \rangle = \frac{1}{8} B_p \langle B_{ph} \rangle \quad (5)$$

On the basis of equation (1), we would expect $\Delta\nu$ to be proportional to the mean square chromospheric field and hence, by equation (5), proportional to the mean absolute photospheric field. This is in accord with the results of Woodard *et al.*¹⁴. If we assume $\langle B_{ph} \rangle$ has a maximum value¹³ of ~ 20 G and B_p is taken to be 1,500 G, then $\langle B_c^2 \rangle^{1/2}$ must be less than ~ 60 G.

The observational evidence thus far suggests that solar-cycle variability of p-mode frequencies originates in the latitudinal bands of magnetic activity^{1,16}. It seems very likely, from the strong correlation between frequency shift and magnetic activity¹⁴, that the crucial perturbations are themselves magnetic. Possible sources for these magnetic perturbations are the fibril field structures of the photosphere and the canopy field in the overlying chromosphere. On the basis of the calculations done so far, both sources give good agreement with the observed form of the frequency shifts at low frequencies. It seems certain that

the structure of the upper chromospheric perturbations is required to explain the high-frequency effects. $\square$

Received 18 July; accepted 22 October 1991.

1. Libbrecht, K. G. & Woodard, M. F. *Nature* **345**, 779–782 (1990).
2. Eisworth, Y., Howe, R., Isaak, G. R., McLeod, C. P. & New, R. *Nature* **345**, 322–324 (1990).
3. Campbell, W. R. & Roberts, B. *Astrophys. J.* **338**, 538–556 (1989).
4. Evans, D. J. & Roberts, B. *Astrophys. J.* **356**, 704–719 (1990).
5. Evans, D. J. & Roberts, B. *Astrophys. J.* **371**, 387–395 (1991).
6. Duvall, T. L. *Nature* **300**, 242 (1982).
7. Jones, H. P. & Giovanelli, R. G. *Solar Phys.* **87**, 37–42 (1983).
8. Spruit, H. C. *Solar Phys.* **34**, 277–290 (1974).
9. Gough, D. O. in *Proc. Progress of Seismology of the Sun and Stars, Lecture Notes in Physics* **367** (eds Osaki, Y. & Shibahashi, H.) 283–318 (Springer, Berlin, 1990).
10. Goldreich, P., Murray, N., Willette, G. & Kumar, P. *Astrophys. J.* **370**, 752–762 (1991).
11. Gough, D. O. & Thompson, M. J. in *Proc. Advances in Helio- and Asteroseismology, IAU Symp.* **123** (eds Christensen-Dalsgaard, J. & Frandsen, S.) 175–180 (Reidel, Dordrecht, 1986).
12. Dziembowski, W. A. & Goode, P. R. *Astrophys. J.* **376**, 782–786 (1991).
13. Spruit, H. C. & Roberts, B. *Nature* **304**, 401–406 (1983).
14. Woodard, M. F., Kuhn, J. R., Murray, N. & Libbrecht, K. G. *Astrophys. J.* **373**, L81–84 (1991).
15. Spruit, H. C. in *The Sun as a Star*, NASA SP-450 (ed. Jordan, S.) 385–412 (NASA, Washington, D.C. 1981).
16. Kuhn, J. R. *Astrophys. J.* **331**, L131 (1988).

ACKNOWLEDGEMENTS. We thank P. Goode, M. Schussler, T. Tarbell and M. Thompson for discussions and comments. We acknowledge financial support from the SERC.

Distribution of plasma turbulence in our Galaxy derived from radio scintillation maps

Chris Argyrou Hajivassiliou

Mullard Radio Astronomy Observatory, Cavendish Laboratory,
Madingley Road, Cambridge CB3 0HE, UK

THE scattering of radio waves by plasma turbulence in the Galaxy is analogous to the scintillation of starlight due to the Earth's atmosphere, and similarly leads to a broadening of the apparent angular sizes of extragalactic radio sources^{1–3}, as well as the smearing out of pulsar pulse profiles. Previous investigations of the galactic plasma turbulence using these effects have suggested a monotonic increase in plasma turbulence towards the galactic plane^{4–6} and towards the centre^{7,8}. Here I present a new map of galactic plasma turbulence derived from the results of a recent interplanetary scintillation radio survey⁹, using a method of analysis based on the angular size distribution of radio sources. The map reveals that structure persists to high galactic latitudes, and that it agrees with the morphology of the soft X-ray background. These results suggest that the solar neighbourhood is encapsulated in an envelope of plasma turbulence, most probably the relic of the supernova explosion of a nearby massive star.

The strength of the galactic plasma turbulence in any region of the sky is reflected in a collective increase in the apparent angular sizes, Θ , of radio sources in that direction. New theory of the interplanetary scintillation survey process^{10,11} shows that a sensitive indicator of this increase is the probability, $P_c(\Theta < \Theta_c)$, that the angular size is less than a threshold value, Θ_c , here 0.32", set by maximizing the variation in P_c with respect to change in the level of scattering. A consistent estimator of P_c is the fraction, f , of sources with $\Theta < 0.32''$ in a given patch of sky.

The map of the variation of f with galactic coordinates is shown in Fig. 1. As expected from previous work, there is evidence for increased scattering towards the Galactic Centre⁷ and close to the plane⁵, but the well-defined structure at $l \approx 90^\circ$ and $l \approx 270^\circ$, which persists to high galactic latitudes, has not been noted before either in interplanetary scintillation observations or in observations of pulsars. The new regions of enhanced scattering do, however, correspond well with regions of enhanced soft X-ray emission¹², interpreted in terms of hot gas filling a local bubble which encloses the solar neighbourhood¹³. The X-ray data suggest¹⁴ that this bubble is elongated perpendicular to the galactic plane and extends to ~ 150 pc away from

FIG. 1 Map of the variation of f , measured in bins $8^\circ \times 16^\circ$ ($b \times l$), plotted as an Ait-off equal-area projection of the sky in new²⁷ galactic coordinates, with the centre of the Galaxy at the centre of the plot and the north polar spur to the left at $l \approx 30^\circ$. Values of f have been calculated by applying a new method of analysis¹⁰ to the Cambridge interplanetary scintillation survey⁸ at 81.5 MHz. On average, eight sources were used for each independent calculation of f , making the typical associated 1σ error ~ 0.05 . In this colour representation, f ranges from 0 (red: maximum turbulence) to 0.25 (black: minimum turbulence). The large grey area centred on $l \approx 300^\circ$, $b \approx -30^\circ$ is the part of sky lying below the Cambridge horizon which was not covered in the survey.

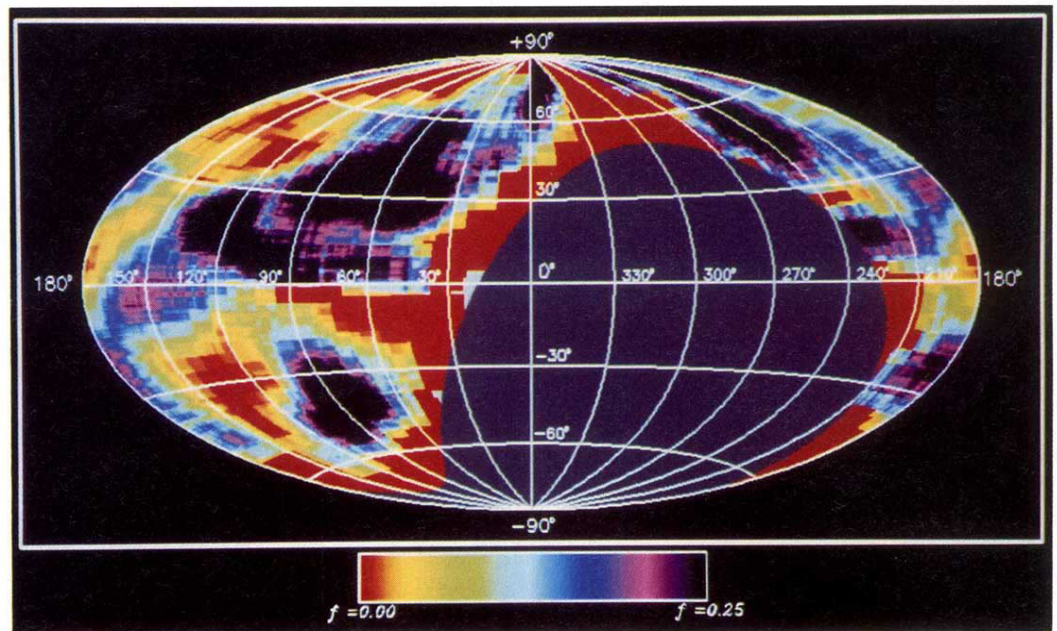
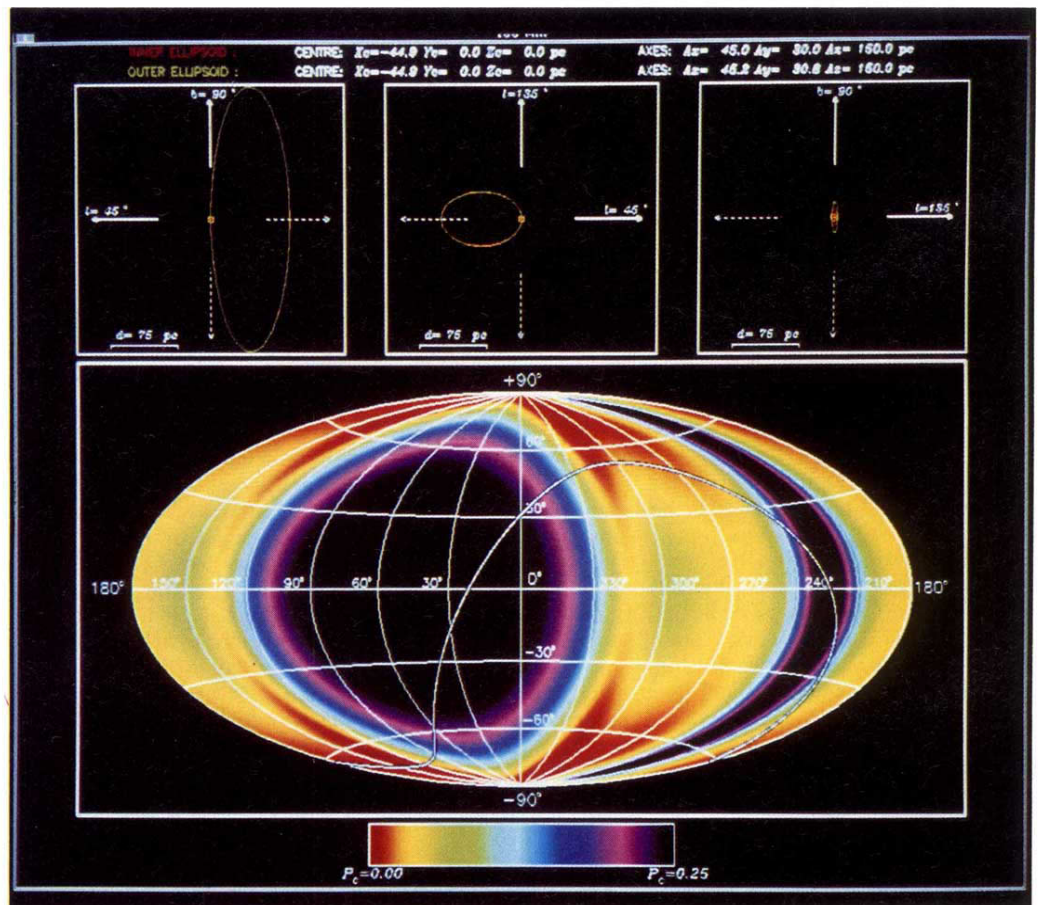


FIG. 2 Map of P_c expected on the basis of a model of the local interstellar medium in which the solar neighbourhood is encapsulated by an ellipsoidal envelope of plasma turbulence, ~ 0.8 pc in thickness, 150 pc in extent perpendicular to the plane of the Galaxy and having an elliptical cross-section in the plane with semimajor axis on the $l=45^\circ$, 225° great circle, 45 pc in length and semiminor axis 30 pc (dimensions shown at the top). In the boxes (from left to right) are shown cuts through the solar location (1) perpendicular to the galactic plane in the direction $l=45^\circ$, (2) in the plane, and (3) perpendicular to it in the direction $l=135^\circ$. The level of broadening when the line of sight cuts through the shell (assumed to contain plasma turbulence with Kolmogorov²⁸ spectrum of strength $C_N^2 = 0.7 \text{ m}^{-20/3}$) is first calculated² and then used to determine P_c from the theory^{10,11} of the interplanetary scintillation survey. This value is plotted using the same projection and colour representation as in Fig. 1; the curve running across the map marks the Cambridge horizon. Enhanced broadening occurs in directions tangential to the shell, leaving two 'holes' of minimum broadening (black) in directions perpendicular to it. We are assumed to be very close to the shell in the direction $l \approx 45^\circ$, and for this reason the hole



a Cambridge horizon. Enhanced broadening

covers a large area of sky. The crescent-shaped region of minimum broaden-

ing covers a large area of sky. The crescent-shaped region of minimum broadening, centred on $l=225^\circ$, corresponds to the other hole, which is farther away and therefore subtends a smaller angle, as suggested by the X-ray data and Fig. 1.

it; it is thus much more extended out of than in the plane, where a nominal radius of ~ 45 pc is representative. Sections cut parallel to the plane are roughly elliptical with the semimajor axis on the $l = 45^\circ, 225^\circ$ great circle. The Sun lies at the fringes of the envelope in the direction $l = 30^\circ$.

Interstellar scattering traces regions of high energy input into the interstellar medium, such as shock fronts associated with both plasma turbulence and radiation emission. I therefore postulate that the large-scale features in Fig. 1 originate in highly turbulent, shocked plasma in the envelope of the X-ray bubble. That the wavelength of the radiation emitted from the shocked region should fall in the soft X-ray domain is purely coincidental but, nevertheless, is consistent with a relatively energetic shock which would in turn make observable interstellar scattering all the more probable. I test this hypothesis in Fig. 2, which shows the map of P_c expected if all the scattering originated in an ellipsoidal shell (shown as an inset) whose dimensions and morphology follow those of the X-ray bubble. The directional correspondence between the two maps is remarkable (especially in view of the simplicity of the model and the fact that no attempt has been made to improve the fit by deviating from ellipticity). Subtracting one from the other leaves a map that is essentially featureless over most of the sky, but nevertheless highlights the disagreement in regions close to the plane and towards the Galactic Centre, where enhanced broadening associated with the galactic disk is to be expected. When such regions are excluded, this model can account for the enhanced broadening in the directions $l \approx 135^\circ$ and 315° and towards high latitudes, for the large 'hole' centred around $l \approx 45^\circ$ (cut by the north polar spur to the north) and for the smaller crescent centred on $l \approx 225^\circ$ (of which only the northern part is visible in Fig. 1). The last of these is a sensitive indicator of the elliptical shape predicted by the X-ray data and cannot be reproduced by a spherical model for the bubble.

It is conceivable that the directional anisotropy in f could be intrinsic to the extragalactic sources themselves. This, however, would be difficult to envisage, even in a rotating universe, and the local interpretation is much the more probable (especially in view of the close correspondence between the observed anisotropy and that expected on the basis of the inferred morphology and orientation of the X-ray bubble¹⁴). The proposed scenario therefore explains why the observed number of compact extragalactic radio sources appeared to be unexpectedly low, a fact previously thought to have cosmological significance¹⁵, and also accounts for localized depletions of such sources, previously taken as indicative of large voids ($\sim 1,000$ Mpc) in the matter distribution of the Universe¹⁶. It also suggests that the anisotropic component of the observed soft X-ray background is local in origin but has, nevertheless, implications far beyond how free electron density is distributed in the Galaxy. The hot gas inside the bubble would contain the expansion of the heliosphere and therefore account for the proximity^{17,18} of the heliopause, if this does finally turn out to be as close¹⁹ as suggested in ref. 20. The dense envelope would influence the cosmic-ray transport and confinement in the solar neighbourhood. Its effects, especially in directions tangential to it, should also be taken into account when using dispersion measures of pulsars to deduce their distances. This consideration would limit the significance of the observational test adopted by Sciama²¹ in support of the decaying dark matter hypothesis, reinforcing the negative result of a recent attempt to detect the 14-eV emission line²² that this theory predicts.

Such an X-ray emitting cavity could have been created by the supernova explosion of a single massive progenitor¹³, or have evolved in the fashion of a superbubble^{23–25} by the sequential explosion of many normal stars. As the model shell is so uniform, one would favour the former possibility; the estimated value of $0.7 \text{ m}^{-2/3}$ for C_N^2 , the Kolmogorov parameter describing the strength of the turbulence, can be converted²⁶ to a r.m.s. free electron density $n_e \approx 2 \text{ cm}^{-3}$ suggesting a mass excess in the shell

of ~ 200 solar masses, attributable to the progenitor. This implies a very energetic supernova which must have occurred ~ 45 pc away in the direction $b \approx 0^\circ, l \approx 225^\circ$ (the centre of the ellipsoidal model). No candidate pulsar appears at the predicted location; if a black hole is what has been left behind, its relative proximity and the comparatively clear line of sight should increase our chances of ever detecting one. $\square$

Received 10 June; accepted 5 November 1991.

1. Scheuer, P. A. G. *Nature* **218**, 920–922 (1968).
2. Blandford, R. & Narayan, R. *Mon. Not. R. Astr. Soc.* **213**, 591–611 (1985).
3. Rickett, B. J. *Ann. Rev. Astr. Astrophys.* **28**, 561–605 (1990).
4. Readhead, A. C. S. & Hewish, A. *Nature* **236**, 440–443 (1972).
5. Cordes, J. M., Ananthakrishnan, S. & Dennison, B. *Nature* **309**, 689–691 (1984).
6. Cordes, J. M., Weisberg, J. M. & Boriakoff, V. *Astrophys. J.* **288**, 221–247 (1985).
7. Pramesh Rao, A. & Ananthakrishnan, S. *Nature* **312**, 707–711 (1984).
8. Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. Astr. Soc.* **213**, 613–639 (1985).
9. Purvis, A., Tappin, S. J., Rees, W. G., Hewish, A. & Duffett-Smith, P. J. *Mon. Not. R. Astr. Soc.* **229**, 589–619 (1987).
10. Hajivassiliou, C. A. thesis, Univ. of Cambridge (1990).
11. Hajivassiliou, C. A. & Duffett-Smith, P. J. *Mon. Not. R. Astr. Soc.* **247**, 491–499 (1990).
12. McCammon, D., Burrows, D. N., Sanders, W. T. & Kraushaar, W. L. *Astrophys. J.* **269**, 107–135 (1983).
13. Cox, D. P. & Anderson, P. R. *Astrophys. J.* **253**, 268–289 (1982).
14. Cox, D. P. & Snowden, S. L. *Adv. Space Res.* **6**, 97–107 (1986).
15. Hewish, A., Readhead, A. C. S. & Duffett-Smith, P. J. *Nature* **252**, 657–659 (1974).
16. Artyukh, V. S. & Ogannisyan, M. A. *Sov. Astr. Lett.* **14**, 377–378 (1988).
17. Axford, W. I. & Ip, W.-H. *Adv. Space Res.* **6**, 27–32 (1986).
18. Holzer, T. E. *Ann. Rev. Astr. Astrophys.* **27**, 199–234 (1989).
19. Suess, S. T. & Dessler, A. J. *Nature* **317**, 702–703 (1985).
20. Kurth, W. S., Gurnett, D. A., Scarf, F. L. & Poynter, R. L. *Nature* **312**, 27–31 (1984).
21. Sciama, D. W. *Nature* **346**, 40–42 (1990).
22. Davidsen, A. F. et al. *Nature* **351**, 128–130 (1991).
23. Heiles, C. *Astrophys. J.* **229**, 533–544 (1979).
24. Heiles, C. *Astrophys. J. Suppl.* **55**, 585–595 (1984).
25. Tomisaka, K. *Adv. Space Res.* **6**, No. 2, 109–118 (1986).
26. Cordes, J. M., Spangler, S. R., Weisberg, J. M. & Clifton, T. R. *AIP Conf. Proc.* **174**, 180–190, New York (1988).
27. Blaauw, A., Gum, C. S., Pawsey, J. L. & Westerhout, G. *Mon. Not. R. Astr. Soc.* **121**, 123–131 (1960).
28. Armstrong, J. W., Cordes, J. M. & Rickett, B. J. *Nature* **291**, 561–564 (1981).

ACKNOWLEDGEMENTS. I thank A. Hewish for his help and encouragement and P. J. Duffett-Smith for comments and suggestions on the manuscript.

Recent grazing impacts on the Earth recorded in the Rio Cuarto crater field, Argentina

Peter H. Schultz* & Ruben E. Lianza†

* Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

† LTV Aircraft Products Group, PO Box 655907, MS-1502, Dallas, Texas 75265, USA

THE most probable angle of a meteoroid impact on a planet is 45° , and an impact at 15° from the horizontal or lower is as likely as at 75° or higher^{1,2}. Yet little direct evidence for oblique impacts exists on the Earth, for two reasons. Unless the impact angle is very low, any asymmetry created during the initial transfer of energy from impactor to target is lost as the crater is formed³; moreover, the shallow craters formed by oblique impact are more easily obscured by subsequent erosion. During routine flights two years ago, however, one of us (R.E.L.) noticed an anomalous alignment of oblong rimmed depressions (4 km $\times$ 1 km) on the otherwise featureless farmland of the Pampas in Argentina. We argue here, from sample analysis and by analogy with laboratory experiments, that these structures resulted from a low-angle impact and ricochet of a chondritic body originally 150–300 m in diameter.

Ten oblong depressions with 4:1 length-to-width ratios occur along a 50-km line running northeast to southwest just north of the Rio Cuarto River and the city of Rio Cuarto, Argentina (Fig. 1). Eight of these structures occur within a 30-km-long corridor only 2 km wide, centred near $64^\circ 14'$ W and $30^\circ 52'$ S. The largest structure (A, to the north) is 4.5×1.1 km (rim to rim) with a slightly smaller adjacent pair (D and E, each $\sim 3.5 \times 0.7$ km) occurring 11 km to the southwest. Three more adjacent

smaller structures (G, F and H, 0.1–0.3 km wide) occur ~5 km farther southwest of D and E, with a subtle but unmistakable lineated pattern fanning out farther to the southwest. Two possibly related oblong features (I, J) occur 30 km southwest of the northern rim of A but are offset 5 km from the principal axes. The western group consists of an oblong pair connected end-to-end with an axis directed toward crater A.

The rugged terrain surrounding each structure generally prevented farming, thereby preserving much of the terrain from agricultural development. Even less-pronounced structures which were subsequently farmed, however, retain their outline through drainage control and vegetation contrasts. The largest structures (A, D and E) have poorly defined rims at either end of the long axes but well defined rims to either side reaching 3–7 m above the surrounding plains. The floors are ~7 m below the plains, and the larger examples show evidence for crater fill, fed by captured drainage. The smaller structures (C, F, G and H) paradoxically are better defined. The elongate depression (C) resembles a long groove, whereas the broader structure (F) contains a central mound with a lineated texture extending farther to the south. The region is covered by 25–50 m of loess accumulated during the Pleistocene; crystalline outcrops and surrounding metamorphic sequences of the Sierra de Cordoba constitute the closest bedrock exposures ~30 km to the north-west⁴. Streams and rivers along the 50-km-long crater chain show evidence for blockage, diversion and capture (Fig. 1).

Because of the geological setting, these structures could easily be interpreted as aeolian in origin. Considerable evidence exists for aeolian reworking, and other deflation basins exist on the Pampas. The unusual pattern and striking resemblance to oblique clustered impacts in laboratory experiments⁵, however, led us to analyse representative samples and fulgurite-like objects collected from the southern end of crater D (Fig. 3a). Their morphologies range from glassy slabs to dumb-bell shapes. Vesicularity also varies, with loess commonly embedded near the surface and in the vesicle walls. Loess particles within the glass commonly have been partially resorbed into the melt fraction. The colour of the glass ranges from a characteristic lime green to a deep olive-black, generally correlated with decreasing vesiculation. Many samples retain a glossy tachylitic sheen on one side, with banded, twisted and pulled textures; cast-like impressions with a granular texture are often found on the other side. Less-vesiculated splash-like examples contain well-defined layering reflecting distinct contrasts in composition and cooling histories. Thin sections reveal flow patterns, immiscible blebs and small inclusions all characteristic of impactites. Electron-microprobe analysis of samples with abundant vesicles confirm this inferred compositional variation with the darkest regions having an FeO content considerably higher than that characterizing the bulk target loess (Table 1), but perhaps indicating resorption of magnetite grains.

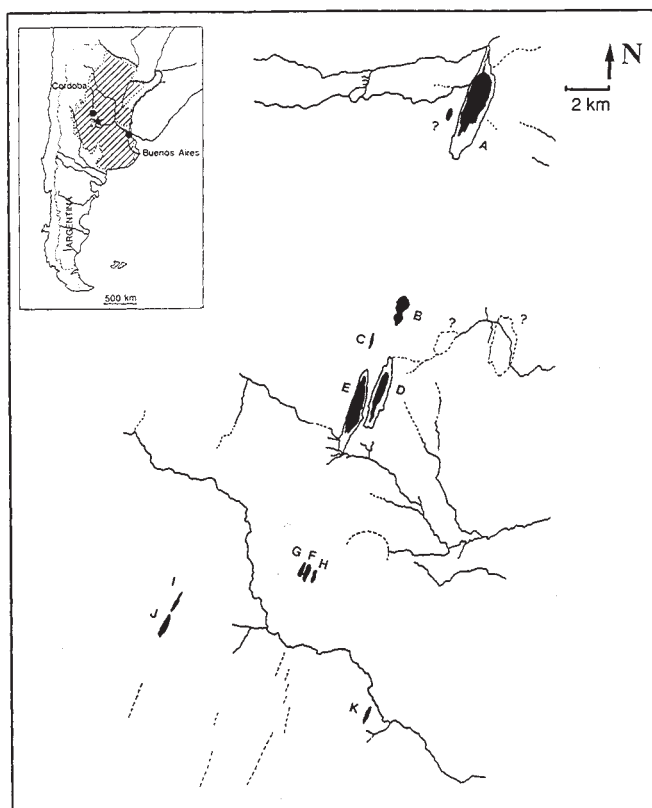


FIG. 1 General sketch map of Rio Cuarto crater field. In contrast to the pattern of smaller craters associated with atmospheric breakup, this group is characterized by a large (1.2 × 4 km), shallow up-range crater (A) with successively smaller craters down range. This type of pattern occurs in laboratory impact experiments and on planetary surfaces for shallow impact angles. The formation of the crater field interrupted, redirected and captured some of the drainage patterns. The dotted lines indicate more subtle drainage signatures. Inset shows location with striped area indicating distribution of loessoid deposits (after ref. 4). Craters A, C, E and D can be seen in the aerial photograph on the cover of this issue of *Nature*.

Among the samples are long, twisted glassy objects (1.5 cm long, 0.5 cm diameter) with a greenish hue, and sections reveal large vesicles along the surface with microvesicles inside. The swirling flow pattern, microvesicles, quartz inclusions and bulk index of refraction (1.512) closely resemble impactites and some tektites found elsewhere on the Earth (see, for example, ref. 6). Microprobe transects across vesicle-poor samples reveal a very homogeneous oxide distribution and high silica content, characteristics in common with tektites⁶. Table 1 further demonstrates

TABLE 1 Major elements from microprobe analysis of selected samples (wt %)

	Chondrite (interstitial glass in melt rim)	Loess*		Clear	Vesicle-abundant impactite†			Darkest brown	Vesicle-poor glass drop‡	HCa- australites§
		Rosario	Miramar		Light brown	Brown	Dark brown			
SiO ₂	50.3	65.5	57.2	58.4	58.3	50.9	55.3	48.6	69.0 (0.5)	68.9–79.7
Al ₂ O ₃	6.8	15.6	17.3	25.1	15.8	18.2	15.3	14.5	16.3 (0.16)	9.3–14.3
MgO	6.3	1.4	1.7	0.34	1.7	2.3	2.8	2.3	1.4 (0.04)	1.31–2.53
Na ₂ O	1.5	3.1	ND	5.4	4.3	3.1	3.7	3.0	3.5 (0.16)	1.00–1.24
K ₂ O	0.27	1.6	ND	3.2	2.8	1.8	2.4	1.8	2.6 (0.08)	2.14–2.21
CaO	4.41	2.1	2.8	6.5	4.15	8.0	4.8	5.0	3.4 (0.11)	1.83–5.48
TiO ₂	0.22	1.1	ND	0.0	2.0	1.4	2.7	3.5	0.64 (0.03)	0.49–0.81
FeO	30.1	6.7	5.4	1.4	10.3	13.7	14.0	20.1	3.8 (0.07)	3.57–4.75
MnO	0.22	—	ND	0.04	0.27	0.20	0.26	0.44	0.10 (0.04)	—

* Cited for different locations in ref. 4.

† Variation within a sample similar to that shown in Fig. 3a.

‡ Average of 5 analyses in transect (standard deviation in %).

§ Ref. 7. High calcium (HCa) tektites.



FIG. 2 One of the smaller members of the Rio Cuarto crater field (crater F, Fig. 1) ~250 m across. The ridge and grooved terrain extending to the southwest is typical of craters formed by a tight cluster of small projectiles in the laboratory.

the general similarity among the high-silica glass, the target loess and high-calcium australites⁷.

Equally intriguing is a small (1×0.5 cm) button-shaped object which is magnetically susceptible (and therefore FeO-rich), found in crater D. Thin sections revealed it to be an H4 chondrite surrounded by a ferrous-oxide-rich zone with carbonate-cemented loess (Fig. 3b). Between this zone and the chondrite is a very well defined zone of re-crystallized melt with blade crystals of olivine. The flange-like sides of the sample resemble the gross morphology of tektites produced by aerodynamic remelting during atmospheric entry⁷. A zone containing acicular olivine occurs on one side and indicates a relatively slow cooling rate. Interstitial glass in the innermost portion of this zone is derived from the chondrite, and microprobe analysis indicates a high concentration of iron oxide (Table 1). This sample and another recovered chondrite fragment indicate the only known direct physical associations between a chondritic meteorite and a large impact crater.

Impactites at Rio Cuarto are not widely exposed, but embedded loess and similarities among samples from three separate craters (C, D and E) support an impact origin for the entire crater field. With this working hypothesis, the crater chain could have formed either by the break-up of an object before impact, or by ricochet and down-range impacts from a single low-angle collision.

In laboratory experiments³, oblique impacts into granular targets under vacuum conditions typically produce relatively circular craters in plan view until the impact angle falls below $\sim 7^\circ$. Crater elongation at such low angles largely results from down-range impacts by ricocheted debris which is no longer consumed by crater growth^{3,8,9} in the milliseconds following impact. Down-range ricochet impacts are clearly evident in strength-controlled cratering experiments, where the ratio of crater to projectile diameter is only 2–4. For gravity-controlled particulate targets at laboratory scales, however, the ratio between crater width and projectile diameter at 6 km s^{-1} exceeds 50 at impact angles of 90° and 30 at 15° (from the horizontal). Scaling considerations^{3,10} indicate that at the size of the Rio Cuarto crater A, this ratio for impact into loess should be less than 9 (impactor at 25 km s^{-1} impact velocity with diameter of 150 m). Consequently, the elongated trench produced (the signature of energy transfer) during oblique entry into the target makes up a greater fraction of the final crater. From scaling considerations, this elongation should increase with size and would be enhanced if the impactor had failed just before impact.

The breached or jagged rims and grooved extensions directed southwest suggest that the Rio Cuarto field was created by an object (or objects) coming from the northeast. Such an interpretation implies that smaller craters occur down range, unlike the pattern of aerodynamic sorting associated with much smaller crater fields created by atmospheric break-up before impact¹⁰. The pattern of impacts at Rio Cuarto is very similar, however, to laboratory experiments and planetary analogues where low angles ($<15^\circ$) result in decapitation of the projectile^{8,9}. Peak pressures and the degree of disruption in laboratory experiments are observed to decrease with impact angle, whereas shear heating increases⁸. As a result, unmelted impactor fragments hit the ground again down range at reduced velocities (50% of the original impact velocity) and at distances dependent on impact angle and target composition. In this picture, Rio Cuarto crater A would represent first contact with a relatively intact impactor, whereas most of the remaining craters were formed by ricocheted fragments striking the ground down range. The low-relief rim of the larger structures and the paucity of large blocks from the basement below the loess are consistent with low-angle strikes.

The proposed sequence of events would help to account for recovered impactor fragments which survived impact by being gradually decelerated in an interacting vapour cloud rather than suffering disruptive shock rarefaction. At impact, large decapitated portions of the object spalled and sheared to a fraction (10–50%) of its original velocity with smaller fragments entrained in the trailing wake. Such survivors of the initial collision should show little evidence for shock effects and partial melting along interfaces. Larger decapitated fragments struck down range at significant fractions of the initial impactor velocity, and the ensuing vapour cloud created a counterflow, thereby decelerating trailing impactor fragments and impactites more slowly. Considerable melting and vaporization of the target should occur even without evidence for high shock levels in the surviving pieces of the impactor.

The Rio Cuarto crater field provides a rare opportunity to investigate energy partitioning associated with large-scale oblique impacts and to explore the complex interactions between impactor, target and atmosphere previously only observed in the laboratory and preserved in different planetary settings^{8,9}. The size of crater A indicates that the impact body was about 150 m in diameter, resulting in an estimated energy release of ~ 350 megatons (TNT equivalent), 30 times larger than the 1908 Tunguska event (for an assumed impact velocity of 25 km s^{-1}). The age of the emplaced loess, the excellent preservation state

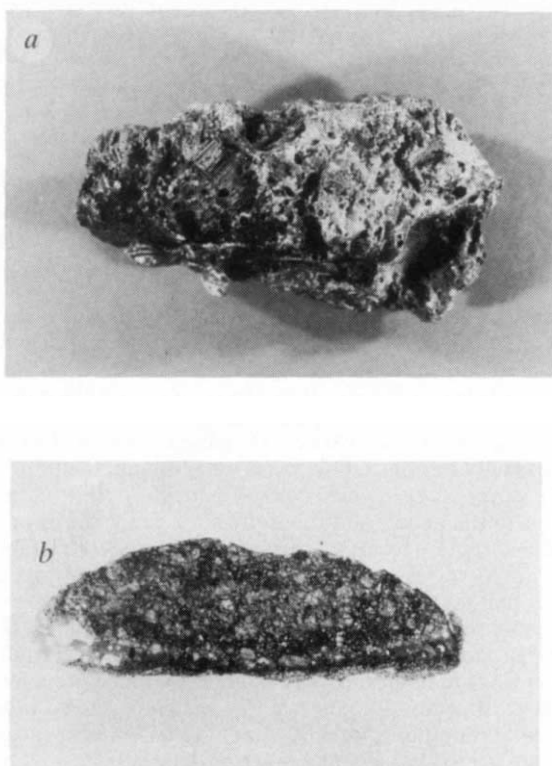


FIG. 3 *a*, Typical vesicular impactite found in crater D (Fig. 1). Tachylitic sheen is typically preserved on one side, thereby indicating a relatively young age. Sample is 10 cm long. *b*, Thin section (transmitted light) of a (2 × 6 cm) button-shaped H4 chondrite found in crater D. A flange-like rim surrounds the flattened surface with a layer of fragmental material (largely loess) adhered to the surface. A 0.5-mm-wide zone of bladed crystals of olivine occurs between the adhered fragmental material and the preserved chondrite. This acicular zone makes a razor-sharp contact with the meteorite and has severed a large chondrule. The development of these crystals indicates a slower cooling history than that characteristic of other recovered impact products. This fragment may represent a spalled or sheared portion of the impactor that formed crater A, which accompanied the larger body (or bodies) producing crater D.

of the smaller craters, the preserved glassy sheen on the impactite and the state of the recovered chondrite all suggest an age considerably less than 10,000 years. Consequently, it is conceivable that this event was witnessed by early inhabitants of South America. □

Crushing C₆₀ to diamond at room temperature

Manuel Núñez Regueiro^{*†}, Pierre Monceau^{*}
& Jean-Louis Hodeau[‡]

^{*} Centre de Recherches sur les Très Basses Températures, CNRS, BP 166X Cedex 9, 38042 Grenoble, France

[‡] Laboratoire de Cristallographie, CNRS, BP 166X Cedex 9, 38042 Grenoble, France

[†] On leave of absence from CNEA and CONICET, Centro Atómico Bariloche, 8400 Bariloche, Argentina

C₆₀ MOLECULES are extremely stable, withstanding hydrostatic pressures of up to at least 20 GPa (ref. 1). It has been proposed that at high pressures they could form a solid harder than diamond². On the other hand, electrical resistivity measurements³ have revealed the formation of an insulating phase above 20 GPa, which was attributed to the low-symmetry state found in X-ray diffraction studies under nonhydrostatic compression¹. Here we report that rapid, nonhydrostatic compression of C₆₀ to pressures of 20 ± 5 GPa transforms it instantaneously into bulk polycrystalline diamond at room temperature. Our measurements place a limit on the stability of the C₆₀ molecular phase under nonhydrostatic pressure. The high efficiency and fast kinetics at room temperature suggest the possibility of using this transformation for fabrication of industrial diamonds.

Fullerenes, cage-like carbon structures made up of twelve pentagons and an even number of hexagons, have only recently been produced in macroscopic quantities⁴. The most prominent member of the family, the football-shaped C₆₀, has interesting and practical properties. Doped with alkalis⁵, it becomes superconducting at 33 K. Its high-pressure properties are also outstanding, with the possibility of phases harder than diamond being produced, either solely under pressure² or in the form of a metastable phase⁶ with a large amount of sp³ hybridization, similar to diamond. There is in fact a transition to a low-symmetry phase under nonhydrostatic (critical pressure $P_c \approx 15$ GPa)¹ and to an insulator under quasi-hydrostatic pressures ($P_c \approx 20$ GPa)³.

It was impossible in our previous experiments³ to recover the sample for analysis, because the cell disintegrated as pressure was released, at room temperature or even at ~10 K. To maximize the possibilities of recovering part of the sample, we did experiments without a pressure medium, using only the pyrophyllite gasket of 1 mm internal diameter filled with C₆₀ powder. Sample A was made from purified C₆₀ powder obtained by conventional methods and separated by column chromatography provided by P. Bernier and A. Rassat. Full details are provided in ref. 3. Samples B and C were made from commercial (Strem Chemicals) C₆₀ powder with 2–12% of C₇₀, according to the specifications, used in its as-received condition, without any further treatment. A rubber exterior ring was added to retain the fragments of the disintegration presumed to occur on pressure release. For sample A, the diamond anvils were

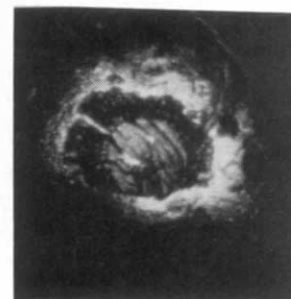


FIG. 1 Photograph of sample A after transformation from C₆₀ to diamond.

Received 21 August; accepted 17 December 1991.

1. Gilbert, G. K. *Bull. phil. Soc. Washington* **12**, 241–292 (1893).
2. Shoemaker, E. in *Physics and Astronomy of the Moon* (ed. Kopal, A.) 283–351 (Academic, New York, 1962).
3. Gault, D. E. & Wedekind, J. A. *Proc. 9th Lunar planet. Sci. Conf. New York*, 3843–3875 (Pergamon, New York, 1978).
4. Teruggi, M. E. *J. sedim. Petrol.* **27**, 322–332 (1957).
5. Schultz, P. H. & Gault, D. E. *J. geophys. Res.* **90**, 3701–3732 (1985).
6. Koerber, C. *Ann. Rev. Earth planet. Sci.* **14**, 323–350 (1986).
7. Chapman, D. R. & Scheiber, L. D. *J. geophys. Res.* **74**, 6737–6776 (1969).
8. Schultz, P. H. & Gault, D. E. *Global Catastrophes in Earth History: Interdisciplinary Conf. Impacts, Volcanism and Mass Mortality*, *Geol. Soc. Am. Spec. Pap.* **247**, 239–261 (1990).
9. Schultz, P. H. & Gault, D. E. (abstr.) *Meteoritics* (in the press).
10. Schmidt, R. M. & Housen, K. R. *Int. J. Impact Engng* **5**, 543–560 (1987).
11. Passey, Q. R. & Melosh, H. J. *J.* **42**, 211–233 (1980).
12. Glass, B. P. *Introduction to Planetary Geology* 145–172 (Cambridge University Press, 1982).

ACKNOWLEDGEMENTS. We thank C. Koerber, R. A. Grieve and H. J. Melosh for reviews and T. Bunch for his preliminary classification of the chondrite. This study was supported by the NSF and NASA.

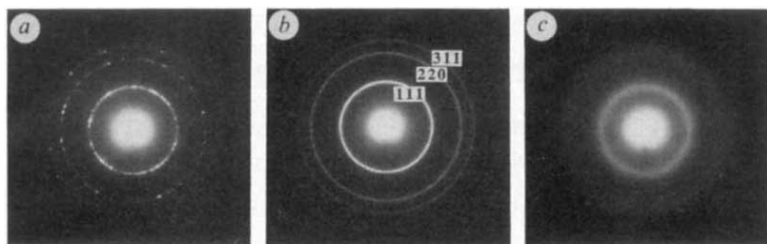


FIG. 2 Typical electron diffraction patterns of chips of the samples. *a*, Well crystallized diamond grains; *b*, small crystallites and *c*, diffuse rings corresponding to 'amorphous', possibly sp^3 , carbon.

slightly slanted, producing a considerable pressure gradient across the cell. For sample C the conditions were the standard ones, whereas for sample B we introduced a fragment of carbon steel razor blade to increase the nonhydrostaticity in the C_{60} powder. We increased pressure at 1 GPa min^{-1} at room temperature, with a strain gauge on the body of the press serving as an approximate manometer. At $25 \pm 5 \text{ GPa}$ for sample A ($20 \pm 5 \text{ GPa}$ for the other samples) there was an explosion similar to those seen previously³. Despite this, we were able to recover the entire samples. Our anvils are made of small diamond crystals ($>40 \mu\text{m}$) sintered with cobalt. When untreated, they are irregular and black. The surfaces have been polished thoroughly, being flat and brilliant, and are clearly discernible from the raw material. After each experiment there was no noticeable change in the texture of the diamond surface in contact with the sample, which conserved its polished aspect.

The samples were still surrounded by their pyrophyllite ring and had a disk shape of diameter 1 mm and height 0.1 mm. They were cracked and could be broken, giving rise to faces with sharp dendrites. Figure 1 shows a photograph of sample A. The samples became transparent, with an amber or reddish-brown colour. There was a certain amount of black powder on the surface and in the pyrophyllite sample boundary. We characterized the sample by X-ray and electron diffraction. Guinier powder patterns of sample A, taken with Fe $K\alpha$ radiation, showed only reflections characteristic of face-centred-cubic (f.c.c.) diamond ($d = 2.06 \text{ \AA}$). X-ray powder patterns of sample C contained no lines and those of sample B showed only a weak broad line at $d = 2.06 \text{ \AA}$ followed by a broader and weaker line at $2.04 \text{ \AA}$. No traces of other phases, such as any C_{60} or graphite material were detected within the precision of this method. We estimate, however, that $\sim 5\text{--}10\%$ of untransformed C_{60} black powder remained.

A small piece of each disk was crushed, scratching the agate mortar. These microscopic chips were transferred to a copper grid coated with holey carbon film, and analysed by electron diffraction. As the grains could be of different carbon phases, pyrophyllite residuals or SiO_2 pieces from the mortar, energy-dispersive X-ray analysis was done directly on each small chip studied by electron diffraction. Typical electron diffraction patterns obtained for carbon chips are presented in Fig. 2. Well crystallized grains with all f.c.c. diamond reflections could be

obtained on samples A and B (Fig. 2*a*); all three samples produced small crystallites which yielded lines corresponding to all diamond reflections (Fig. 2*b*) and 'amorphous' carbon diffuse rings yielded. It is difficult to confirm that the pattern in Fig. 2*c* corresponds to sp^3 -type carbon glass rather than to sp^2 -type, as the lines are very broad and our electron diffraction patterns could be affected by preferential orientation [001]. But sample morphology, as observed by optical or electron microscopy, is not lamellar, and the diffraction patterns evolve continuously evolution from patterns 2*a* to 2*c*. Moreover, all samples were transparent as in sp^3 amorphous carbon. From the peak width at half maximum and from the dark-field images, we can deduce that the size of the crystallites varies from 20 to 1,000 $\text{\AA}$ in the different chips. The largest proportion of well crystallized material was found in sample A where the applied pressure was the least hydrostatic. In contrast, for sample C, made in more hydrostatic conditions, no X-ray lines are observed because the reflections are too broad and weak to be detected. The intermediate case of sample B yields small diamond grains (see the dark-field image corresponding to the diffraction pattern of Fig. 2*b*).

Although it is clear that under hydrostatic conditions C_{60} is stable in its ambient pressure phase up to 20 GPa (ref. 1), we can deduce from our experiments that under nonhydrostatic high pressures C_{60} is unstable to collapse into a diamond phase, even at room temperature.

We note that the sensitivity of the X-rays in detecting carbon compounds is low; the main part of the sample (1 mm diameter by $100 \mu\text{m}$ high) must therefore be in the diamond phase, not merely an optically undetectable chip coming from the anvils. If the diamond sample came from the body of the anvil, we should have recovered a rough, unpolished surface, but we do not. The sample is well surrounded by a pyrophyllite gasket. If the diamond came from the anvil we should have to suppose (unreasonably) that the diamond has squeezed into the pyrophyllite. We also remark that it is on the surface of the samples, which is in direct contact with the anvils, that black powder is found, and that the best crystallites are found in the heart of the sample.

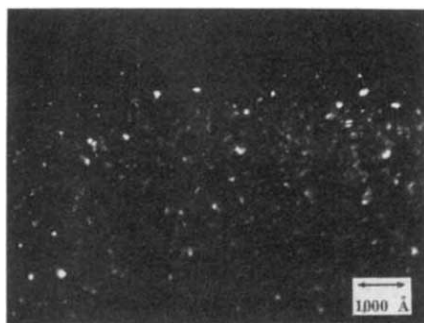


FIG. 3 Dark-field micrograph corresponding to electron diffraction patterns of Fig. 2*b*.

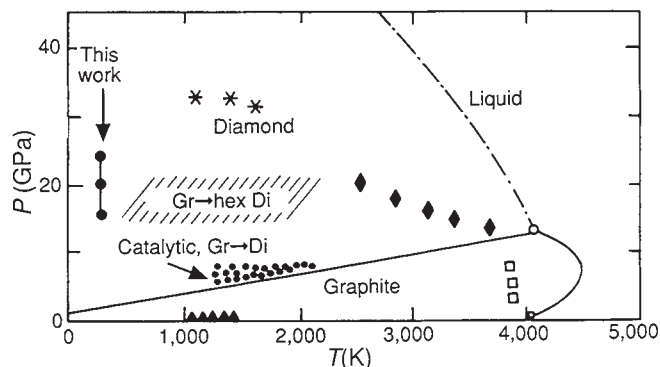


FIG. 4 Carbon phase diagram according to Bundy⁷. □, Region of fast reaction of diamond to graphite; ◆, region of fast reaction, graphite to diamond; *, region of shock conversion of graphite to diamond; ▲, vapour-deposited diamond films; ●, our data.

Figure 4 shows the phase diagram of carbon according to Bundy⁷. Although diamond is the stable phase at room temperature and 2 GPa, it is not possible to produce it under normal conditions because of the extremely low kinetics of the process. Bulk diamond crystals can be made to grow from graphite only when the temperature is raised at least several hundred degrees above room temperature either under shock treatment ($P > 30$ GPa, $T > 1,100$ K)⁸ or in the presence of a catalyst such as molten nickel ($P > 6$ GPa, $T > 1,200$ K)⁹. The difficulty comes from the high stability of the graphitic planes, which melt only above 4,000 K. The transformation from this planar sp^2 structure to the diamond sp^3 network is difficult, and extreme temperatures and pressures are necessary to induce the transformation. The role of the catalyst in the present industrial processes is, in part, to dissolve carbon from graphite, which then precipitates to form the allotrope stable at that temperature and pressure (diamond). Although C_{60} can be considered as a folded graphite sheet, we must take into account the fact that in the pentagons¹⁰ the predominant hybridization is sp^3 . This may make the transformation of C_{60} into diamond easier. A dense assembly of C_{60} spheroids⁶, where 48 out of 60 carbon atoms have quasi-tetrahedral coordination, is sterically fairly close to that of diamond, implying that a small rearrangement of the atoms can result in the change of structure. The fact that we need nonhydrostatic conditions to trigger the reaction may be due to an instability of C_{60} molecules to uniaxial or shear stresses. Fullerenes are probably very stable to isotropic stress, resulting in homogeneous deformation¹, but they possibly cannot resist being squeezed anisotropically. □

Received 28 November; accepted 17 December 1991.

1. Duclos, S. J., Brister, K., Haddon, R. C., Kortan, A. R. & Thiel, F. A. *Nature* **351**, 380–382 (1991).
2. Ruoff, R. S. & Ruoff, A. L. *Nature* **350**, 663–664 (1991).
3. Núñez Regueiro, M., Monceau, P., Rassat, A., Bernier, P. & Zahab, A. *Nature* **354**, 289–291 (1991).
4. Kraetschmer, W., Lamb, L. D., Fostiropoulos, L. D. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
5. Fleming, R. M. *et al.* *Nature* **352**, 787–788 (1991).
6. O'Keeffe, M. *Nature* **352**, 674 (1991).
7. Bundy, F. P. *Physica A156*, 169–178 (1989).
8. De Carli, P. S. & Jamieson, J. C. *Science* **133**, 1821–1824 (1961).
9. Bundy, F. P., Hall, H. T., Strong, H. M. & Wentorf, R. H. *Nature* **176**, 51–55 (1955).
10. Saito, S. & Oshiyama, A. *Phys. Rev. Lett.* **66**, 2637 (1991).

ACKNOWLEDGEMENTS. We thank J.-J. Capponi, J. M. Mignot and D. Núñez-Regueiro for advice and discussions.

Electron paramagnetic resonance studies of lanthanum-containing C_{82}

Robert D. Johnson, Mattanjah S. de Vries,
Jesse Salem, Donald S. Bethune
& Costantino S. Yannoni

Almaden Research Center, IBM Research Division, 650 Harry Road,
San Jose, California 95120-6099 USA

THE conjecture that atoms can be trapped inside closed carbon cages such as the fullerenes was first made by Kroto *et al.*¹. Mass spectroscopic evidence obtained soon after² suggested that lanthanum atoms were encapsulated in fullerenes prepared by laser vaporization of a lanthanum-impregnated graphite disk, and these results were later corroborated^{3,4}. Recently, helium atoms have been incorporated into fullerenes through collisions in the gas phase⁵, and evidence has been obtained for the formation of metal-containing fullerenes during arc burning of composite graphite rods⁶. All of these studies, however, have produced quantities too small for characterization using standard spectroscopic techniques. We report here the preparation of milligram quantities of lanthanum-containing C_{82} , which can be solvent-extracted in yields of about 2% along with empty C_{60} and C_{70} cages. We have

measured the electron paramagnetic resonance spectrum of this mixture, both in solution and in the solid state, which reveals that the lanthanum atom has a formal charge of 3+, and the C_{82} a charge of 3–. This runs contrary to some expectations that the doubly charged fulleride anions would be the most stable species^{6,7}; it also reveals that the fullerene cages have the same formal charge as in the superconducting alkali-metal-doped phases^{8,9}.

Our samples were prepared by using a composite rod made of graphite and La_2O_3 . The materials were mixed with a binder (dextrin), partially dried and pressed into a rod. After annealing at 150 °C, the rod was heated to 1,400 °C for 2 h. The soot produced by arc vaporization of the rod under helium at 200 Torr was extracted with toluene, and the extract was washed with diethyl ether and dried. The resulting powder was analysed using mass spectrometry and electron paramagnetic resonance (EPR) spectroscopy.

The time-of-flight mass spectrum of material laser-desorbed from the dry powder (Fig. 1) shows large peaks for C_{60} , C_{70} and $La@C_{82}$. The inset in Fig. 1 shows the fragmentation pattern found for $La@C_{82}$ when the intensity of the ionization laser (wavelength 193 nm) was increased. This pattern is consistent with the pairwise removal of carbons from the skeletal cage, with the La atom remaining inside the resulting smaller fullerene³. The two groups of peaks centred about 15 mass units above the C_{60} and C_{70} peaks are due to 25% ^{13}C -enriched C_{60} and C_{70} left over in the arc chamber from an earlier burn and inadvertently vacuumed up along with the material produced for this experiment.

The X-band (9.112 GHz) EPR spectrum of the degassed dry powder shown in Fig. 2a is centred at $g = 2.001$, with a hint of rich structure and an overall width of ~10 gauss. No other EPR peaks were detected for 2,500 gauss either side of the field corresponding to $g = 2$. The intensity of the EPR signal shows that the $La@C_{82}$ accounts for $2 \pm 1\%$ (by weight) of the sample, close to the 3.5% estimated from X-ray photoelectron spectroscopy on the same material. The powder was then dissolved in 1,1,2,2-tetrachloroethane, degassed several times using a freeze-thaw cycle and sealed in a 5-mm thin-walled pyrex tube. The solution EPR spectrum, shown in Fig. 2b, consists of eight extremely narrow (0.125 gauss) equally spaced (1.25-gauss interval) lines of equal intensity also centred at $g = 2.0010$, with additional intensity in between. The measured g value is similar to those found for fullerene anion radicals^{10–13}, which range from 1.995 to 2.001. The main eight-line spectrum is unambiguously diagnostic for isotropic hyperfine coupling to a nuclear magnetic moment with spin 7/2 (ref. 14), the value for the ^{139}La nucleus, showing that the unpaired electron has spin

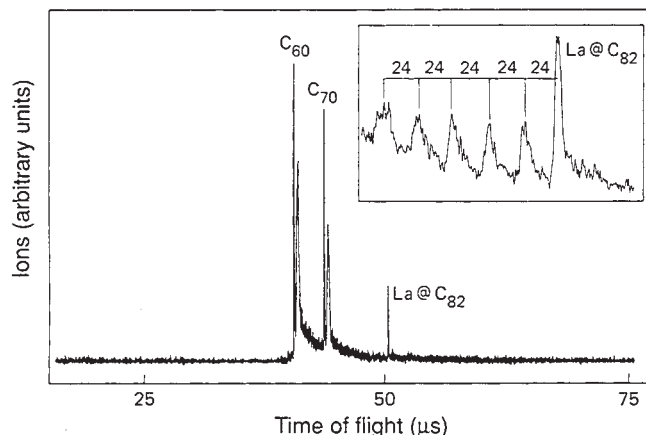


FIG. 1 Mass spectrum of material laser-desorbed from the toluene extract (dried) resulting from arc burning of a composite rod of graphite and La_2O_3 (see text). The inset shows how the peak due to $La@C_{82}$ loses carbons two at a time under strong laser irradiation. The abscissa in the inset is time but the progressive change in mass from peak to peak is shown explicitly.

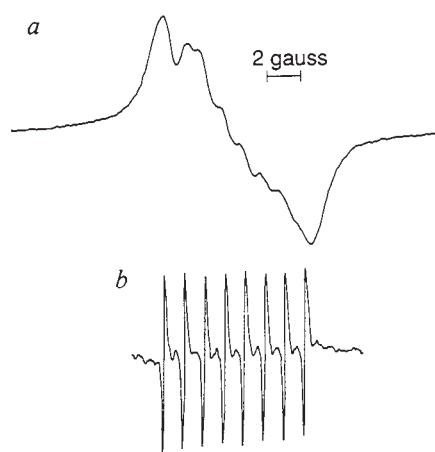


FIG. 2 EPR spectrum (9.112 GHz) at ambient temperature of (a) solid degassed toluene extract (dried) resulting from arc burning of a composite graphite/La₂O₃ rod and (b) a degassed solution of the dried extract in 1,1,2,2-tetrachloroethane.

density on the lanthanum atom. But the ¹³⁹La hyperfine coupling constant (1.25 gauss) is very small; in contrast, the coupling measured¹⁵ for a La²⁺ defect generated in a CaF₂ lattice (the only experimental value we could find in the literature) is ~50 gauss at 20 K. Furthermore, a calculation of the hyperfine coupling for La²⁺ was made using spin-polarized unrestricted Hartree-Fock wavefunctions and yields a value of 186 gauss. Therefore the small value of the observed hyperfine coupling constant indicates that the lanthanum atom is in the 3+ oxidation state. This is consistent with calculations of the electronic structure of La@C₆₀, which found a 3+ charge for the lanthanum atom in that species¹⁶. Figure 2b shows that there are intensity maxima between all the La lines. As 60% of the La@C₈₂ moieties will contain at least one ¹³C, this intensity is consistent with hyperfine coupling to ¹³C in natural abundance; the spectrum is expected to be broadened because of the large number of inequivalent carbons in the C₈₂ cage⁷. A key point, however, is that observation of coupled electron, ¹³⁹La and ¹³C spins provides direct evidence for the association of a La atom with the C₈₂ framework.

The EPR spectra reveal a *g* value consistent with a fullerene anion radical and a La atom in the 3+ oxidation state. We are thus led to a picture of the electronic structure of La@C₈₂ in which the La 6s electrons pair in the lowest unoccupied molecular orbital of C₈₂ while the third electron occupies the next higher energy orbital. Although this leaves La in a diamagnetic oxidation state, polarization of the atomic orbitals by the paramagnetic π -electron system of C₈₂³⁻ results in unpaired spin density at the site of the ¹³⁹La nucleus. Therefore, we believe that La³⁺@C₈₂³⁻ is a ground-state doublet, rather than a closed-shell species as proposed previously^{6,7}, and furthermore that the paramagnetism originates in an unpaired electron spin in the π -electron system of the carbon framework.

An interesting aspect of this result is that the captive ion (La³⁺ in this case) is relatively unperturbed because the interaction with the cage electrons is small. Thus, we have in this bulk sample a large number of single isolated ions at high concentration (~2% of the molecules), potentially making many spectroscopic studies possible. Efforts to separate the empty fullerenes from metal-containing C₈₂ by chromatography are currently in progress. □

- Haufler, R. E. *et al.* *J. phys. Chem.* **94**, 8634–8636 (1990).
- Ros, M. M. & Callahan, J. H. *J. phys. Chem.* **95**, 5720–5723 (1991).
- Chai, Y. *et al.* *J. phys. Chem.* **95**, 7564–7568 (1991).
- Manolopoulos, D. E. & Fowler, P. W. *Chem. Phys. Lett.* (submitted).
- Hebard, A. F. *et al.* *Nature* **350**, 600–601 (1991).
- Holczer, K. *et al.* *Science* **252**, 1154–1157 (1991).
- Krusic, P. J., Wasserman, E., Parkinson, B. A., Malone, B. & Holler, E. R. *J. Am. chem. Soc.* **113**, 6274–6275 (1991).
- Allemand, P.-M. *et al.* *J. Am. chem. Soc.* **113**, 2780–2781 (1991).
- Penicaud, A. *et al.* *J. Am. chem. Soc.* **113**, 6698–6700 (1991).
- Kukolich, S. G. & Huffman, D. R. *Chem. Phys. Lett.* **182**, 263–265 (1991).
- Carrington, A. & McLachlan, A. D. *Introduction to Magnetic Resonance*, Ch. 6 (Harper & Row, New York, 1967).
- Pilla, O. & Bill, H. *J. Phys. C: Solid State Phys.* **17**, 3263–3267 (1984).
- Rosen, A. & Waestberg, B. *Z. Phys. D: At. Mol. Clusters* **12**(1–4), 387–390 (1989).

ACKNOWLEDGEMENTS. We thank P. Kasai, R. MacFarlane, H. E. Hunziker, M. Crowder and R. E. Smalley for discussions, P. S. Bagus for the spin-polarized calculation on La²⁺, D. Hird for XPS data and R. D. Kendrick for help with the *g*-value measurement.

Long-lived photoinduced charge separation in a redox system trapped in a sol-gel glass

Anny Slama-Schwok, Michael Ottolenghi & David Avnir

Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

A KEY feature in the mechanism of photosynthesis is the initial storage of a substantial fraction of the light energy in the form of a long-lived radical pair¹. In attempts to mimic this process in artificial systems^{2–4}, impressive progress has been made in increasing the efficiency of the charge-separation reaction between an excited photosensitizer and an appropriate electron acceptor, but prevention of the energy-wasting back-reaction to neutral species still constitutes a major challenge. The back-reaction limits the length of time during which charge separation (and thus energy storage) can be maintained. Here we report exceedingly long-lived (up to a few hours) photoinduced charge separation in an artificial photosynthetic system that does not require a secondary substrate to react with the charged species. Our system uses pyrene (Py*) as the photosensitized electron donor and *N,N'*-dimethyl-4,4'-bipyridinium (methyl viologen, MV²⁺) as the electron acceptor, both immobilized in a porous sol-gel silica glass⁵. The redox reaction is carried out by the mediation of a third mobile charge carrier in the intrapore space⁶. The spatial separation between the donor and acceptor inhibits the back-reaction to produce the long lifetimes of the charge-separated pair.

Energy storage in artificial photochemical systems generally involves the reaction between an excited photosensitizer A* and an appropriate electron acceptor B:



The back-reaction (2), which in homogeneous solutions is usually diffusion-controlled, must be slowed down to allow the radical pair formed in reaction (1) to undergo subsequent (catalytic) processes leading to useful energy-rich products. Long-lived charge separation is generally achieved by carefully designing heterogeneous environments for carrying out the reaction^{2,3}. Radical pairs with lifetimes of minutes have been produced in such systems^{7–11}. For example, Moser *et al.*¹¹ observed a relatively high yield of long-lived charge carriers in an Fe(III)-doped TiO₂ colloidal semiconductor⁴.

The photochemical reaction system used here involves the reduction of immobilized *N,N'*-dimethyl-4,4'-bipyridinium chloride (methyl viologen, MV²⁺) to MV^{•+} by immobilized excited-state pyrene (Py*), through mediation by the mobile charge carrier, *N,N'*-tetramethylene-2,2'-bipyridinium bromide

Received 24 October; accepted 17 December 1991.

- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
- Heath, J. R. *et al.* *J. Am. chem. Soc.* **352**, 7779–7780 (1985).
- Weiss, F. D., Elkind, J. L., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *J. Am. chem. Soc.* **110**, 4464–4465 (1988).

(TV²⁺), according to the following reactions

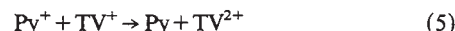


The unidirectionality of the shuttling reaction is assured by the difference in redox potential, E_0 , of the two redox couples MV⁺/MV²⁺ ($E_0 = -0.44$ V) and TV⁺/TV²⁺ ($E_0 = -0.65$ V)¹². About 5% of the initially generated pair (Py⁺, MV⁺) was found to live for at least 4 hours. Py and MV²⁺ were immobilized in a porous silica glass by the sol-gel method⁵, which allows room-temperature trapping of organic molecules in porous inorganic oxide glasses^{6,13-15}. The process involves the polymerization of tetramethoxysilane in the presence of the two reactants, as previously described^{14,15}. We have shown^{6,14,15} that although the trapped molecules cannot be leached out, a significant fraction of them is still exposed to reactions with solutes in the intra-pore liquid phase. Chemical communication between the two trapped molecules is then created by the diffusive shuttling of TV²⁺/TV⁺ in the aqueous pore space. We have shown previously that a charge-shuttling reaction occurred between A^{*} and B, both immobilized in a sol-gel glass, in a different redox system⁶. The reaction of A⁻ with water, however, precluded the determination of the lifetime of the A⁻, B⁺ radical pair.

Figure 1 shows the quenching of glass-entrapped Py^{*} by TV²⁺, present in the aqueous phase which fills the pores of the glass. The quenching rate levels off at high concentrations of TV²⁺, matching the adsorption isotherm of TV²⁺ on the same glass. Similar patterns occur in other redox systems and have been interpreted in terms of both short-range and long-range electron transfer reactions between the intra-pore solute (in this case TV²⁺), adsorbed on the pore surface, and the trapped photosensitizer (Py^{*})¹⁵.

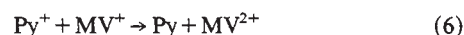
Spectral evidence for the formation of Py⁺ and TV⁺ in the quenching reaction (3), in terms of long-lived products, is presented in Fig. 2a. The difference spectrum recorded 10 s after laser irradiation shows the typical 446-nm peak of Py⁺ and smaller contributions at 365 nm and 490 nm, attributed to TV⁺ (somewhat blue-shifted compared with homogeneous solutions)⁶. The absorption recorded 40 min later (Fig. 2b) is consistent with the production of the (Py⁺, MV⁺) radical pair,

generated in the shuttling reaction (4) in competition with the back-reaction (5):



The increase of MV⁺ absorption, shown by the increase of the peaks at 396 nm and 600 nm, and the decrease of part of the TV⁺ peaks at 365 nm and 490 nm, reflect reaction (4). The 1:1 stoichiometric ratio of MV⁺ and Py⁺ is in keeping with published values of the corresponding extinction coefficients, leading to peak intensity ratios: $\Delta A(396 \text{ nm})/\Delta A(446 \text{ nm}) = 1.3 \pm 0.2$ and $\Delta A(600 \text{ nm})/\Delta A(396 \text{ nm}) = 0.38 \pm 0.05$ (refs 16, 17). These absorbance ratios are suggestive rather than definite because of the uncertainties in the extinction coefficients of the ions in the glassy environment. Figure 2b shows ratios of 1.3 and 0.43 respectively, in agreement with the expected 1:1 stoichiometry. TV²⁺ acts as a shuttler because it is reduced by Py^{*} to TV⁺, which then reduces MV²⁺ into MV⁺ by reaction (4), regenerating the initial TV²⁺. The back-reaction (5) is revealed by the decrease in the intensity of the Py⁺ peak (446 nm), by the regeneration of ground-state Py (336 nm) and by part of the TV⁺ decay (366 and 490 nm). The slow rates of reactions (4) and (5) have been discussed in terms of adsorption and long-range electron transfer phenomena^{6,14,15}.

At [MV²⁺] = 3×10^{-6} M, the characteristic absorbances of MV⁺ and Py⁺ (at 600 nm and 446 nm, respectively), which survive for at least 4 h (Fig. 3), amount to ~5% of the radicals initially formed (as detected 200 ns after the nitrogen laser pulse). This lifetime is shortened as [MV²⁺] is increased (for example to 15 min for [MV²⁺] = 4×10^{-5} M). The concentration dependence can be attributed to the longer shuttling distances of TV⁺ at low [MV²⁺], resulting in larger average spatial separation between Py⁺ and MV⁺, and thus slower recombination rates via the long-range back-reaction (6):



In the absence of TV²⁺, a relatively small yield of the (Py⁺, MV⁺) pair is still observed, but with lifetimes that do not exceed several

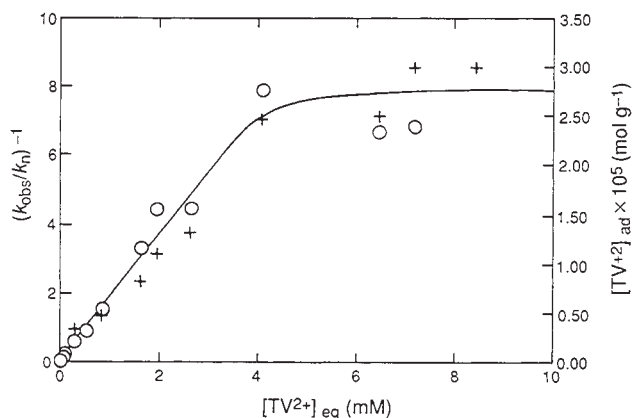


FIG. 1 (+, left scale), Quenching of the fluorescence of glass-trapped Py^{*} as function of the equilibrium concentration of TV²⁺ ([TV²⁺]_{eq}) present in the intra-pore aqueous phase. The Py-doped glass was prepared as described previously^{6,14,15}. The observed Py^{*}-decay rate was analysed according to a three-exponential function. k_{obs} and k_n are the observed decay and natural decay rates of the first exponential component. Similar behaviour was observed for the second component. The third component, unquenched by TV²⁺, is attributed to an unexposed Py population. The calculated quenching rate constant $k_q = (k_{\text{obs}} - k_n)/[\text{TV}^{2+}]_{\text{ads}}$ is $k_q = 4 \times 10^{12} \text{ g mol}^{-1} \text{ s}^{-1}$. TV²⁺ was synthesized as described in ref. 20. (O, right scale), Adsorption isotherm of TV²⁺. The amount of adsorbed TV²⁺ (moles adsorbed per g glass) as a function of [TV²⁺]_{eq} (M).

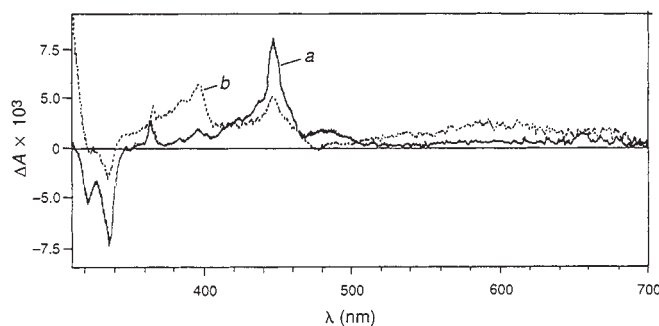


FIG. 2 Difference absorption spectra showing long-lived photoinduced charge separation in the ternary system: Py (4×10^{-5} M) and MV²⁺ (3×10^{-6} M) co-trapped in the sol-gel glass matrix with TV²⁺ ([TV²⁺] = 3.1 mM) present in the a 40-mM NaCl aqueous phase, and de-aerated with 99.999% pure N₂ for 48 h. The glass has a nitrogen-BET surface area of 680 m² g⁻¹ with an average pore size of 20 Å (as determined with a Micromeritics ASAP-2000 surface area analyser). a, Difference spectrum (with respect to the nonirradiated sample) after exposure of the sample to 100 pulses of a N₂ laser (337.1 nm, 3×10^{-4} J, 0.6 ns). The spectrum reveals the generation of Py⁺ and TV⁺. The extinction coefficients of the two TV⁺ peaks are very low, typically 7×10^3 and $3 \times 10^3 \text{ M}^{-1}$ respectively, compared with 3.8×10^4 for Py⁺ (450 nm) and 4.66×10^4 and 1.4×10^4 for MV⁺ at 398 and 600 nm, respectively^{12,16}. The initial amount of MV⁺ present is assigned to faster components of reaction (4). Basically identical changes were observed when we exchanged the pulsed laser source for the 365-nm line of a continuous 100-W mercury lamp. b, Difference spectrum recorded 40 min later. The spectral difference between b and a is attributed to the parallel occurrence of reactions (4) and (5) and is thus in keeping with the equation: $\Delta[\text{TV}^+] = \Delta[\text{Py}^+] - \Delta[\text{MV}^+]$ (where $\Delta[\text{TV}^+] = \Delta A(364 \text{ nm})/\epsilon_{364 \text{ nm}}$, where A is absorbance and ϵ is extinction coefficient. $\Delta[\text{MV}^+] = \Delta A(396 \text{ nm})/\epsilon_{396 \text{ nm}}$; $\Delta[\text{Py}^+] = \Delta A(446 \text{ nm})/\epsilon_{446 \text{ nm}}$).

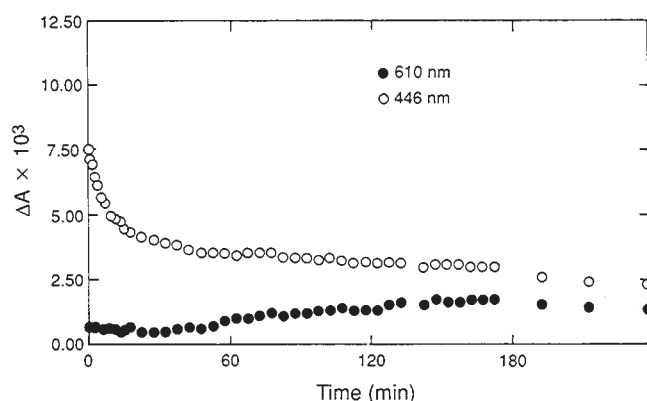


FIG. 3 Time-dependent absorbance changes at wavelengths characteristic of Py^+ (446 nm, $\circ$), and MV^+ (610 nm, $\bullet$), following irradiation of the Py (4×10^{-5} M), TV^{2+} (0.5 mM in 40-mM NaCl), MV^{2+} (3×10^{-6} M) system. Other conditions were as in Fig. 2. The initial decay at 446 nm and increase at 610 nm are due to the superimposed contributions of reactions (4) and (5). The tailing slow decay above ~ 200 min is due to reaction (6).

minutes. This reaction, which most probably involves solvated electrons generated by photoionization of Py^* (ref. 18), is still under investigation. Nevertheless, because side reactions, for example with water, shorten the lifetime of the electron, that process decays and allows the stabler TV^+ to reduce MV^{2+} molecules at much longer distances. We also note that a broad low-intensity peak at 380 nm (attributed to a photo-decomposition product of TV^{2+}) grows over a timescale of several hours without, however, affecting the Py^+ and MV^+ absorptions.

We have thus achieved an extremely long lifetime of the photoinduced charge-separated pair, although in low yield, by immobilizing the primary donor and secondary acceptor in the rigid glass matrix. When treated as a bimolecular process, an upper limit of the rate of charge recombination through reaction (6) yields a second-order rate constant of $k(6) \leq 5 \times 10^2 \text{ mol l}^{-1} \text{ s}^{-1}$ (obtained at $[\text{MV}^{2+}] = 3 \times 10^{-6} \text{ M}$ and at an ionic strength of 40 mM NaCl, the lifetime being independent of the ionic strength). This rate constant is seven orders of magnitude lower than in homogeneous methanolic solutions ($\sim 5 \times 10^9 \text{ mol l}^{-1} \text{ s}^{-1}$)^{15,19}. The retardation of the back-reaction in the present (ternary) trapped system is also at least two orders of magnitude more efficient than that seen¹⁵ in the absence of a shuttler, using trapped Py and adsorbed MV^{2+} . In most previous attempts to achieve long-lived charge separation, there has been some degree of diffusional freedom allowing the back-reaction. This applies to flexible organic polymeric systems, to colloids and to systems based on adsorption^{2,3}. The ability to block the diffusion process in the inert inorganic sol-gel matrix, while still allowing the diffusion of a mediating molecule, is the key to the success of the present system. $\square$

17. Tsukahara, K. & Wilkins, R. G. *J. Am. chem. Soc.* **107**, 2632–2635 (1985).
18. Ottolenghi, M. *Chem. Phys. Lett.* **12**, 339–342 (1971).
19. Das, P. K. *J. Chem. Soc. Faraday Trans.* **79**, 1135–1145 (1983).
20. Homer, R. F. & Tomlinson, T. E. *J. chem. Soc.* 2498 (1960).

ACKNOWLEDGEMENTS. We are grateful to I. Willner, D. Mandler, N. Lapidot and Y. Eichen for a donation of TV^{2+} and for helpful discussions. We thank M. Grätzel for comments. The research was supported by grants from the Israel National Council for Research and Development, the US Army Research, Development and Standardization Group (UK), the E. Berman Research Fund and the Krupp Foundation. Support by the L. Farkas (Minerva) Center for Light-Energy Conversion and by the F. Haber Research Center for Molecular Dynamics is also acknowledged.

A polymer gel with electrically driven motility

Yoshihito Osada, Hidenori Okuzaki & Hirofumi Hori

Department of Chemistry, Ibaraki University, 2-1-1 Bunkyo, Mito, Ibaraki 310, Japan

A SYSTEM capable of converting chemical energy to mechanical energy could serve as an actuator or an 'artificial muscle' in several applications. Here we describe a chemomechanical system of this sort based on a synthetic polymer gel. The gel network is anionic, and positively charged surfactant molecules can therefore bind to its surface, inducing local shrinkage by decreasing the difference in osmotic pressure between the gel interior and the solution outside. By using an electric field to direct surfactant binding selectively to one side of the gel, we can induce contraction and curvature of a strip of gel. Reversing the direction of the field causes contraction of the opposite side, and when the gel is suspended in solution from a ratchet mechanism, it can thereby be made to move with a worm-like motion at a velocity of up to 25 cm min^{-1} .

The isothermal conversion of chemical energy into mechanical work underlies the motility of all living organisms. These systems are efficient because chemical energy is directly converted to mechanical energy without intermediate steps producing heat. Early this century, J. H. van't Hoff showed that chemical free energy could in theory be converted into mechanical work by osmotic cells with pistons. Van't Hoff's theoretical osmotic engine is, of course, very different from living muscles, where the motility is produced by an appropriate assembly of protein molecules.

More than 40 years ago, Katchalsky and collaborators^{1,2} demonstrated that collagen fibres changed dimension reversibly on transition from cyclic helices to random coils when they were immersed cyclically in salt solution and water. Katchalsky *et al.* referred to this as a 'mechanochemical system', defined as a thermodynamic system capable of transforming chemical energy to mechanical work (nowadays, the term 'chemomechanical system' is preferred). Since then, there have been few advances in practical chemomechanical devices, except for some constructional modifications³.

We report here a model of an electrically driven 'muscle', made of water-swollen synthetic polymer gel, which has motility when immersed in water. The system is based on an electrokinetic molecular assembly reaction of surfactant molecules on the hydrogel, which is made of weakly crosslinked poly(2-acrylamido-2-methyl propane) sulphonic acid (PAMPS).

PAMPS gel was prepared by radical polymerization in the presence of 5 mol% crosslinking agent, N,N' -methylene bisacrylamide, as described previously^{4,5}. A strip of PAMPS gel, 1 mm thick, 5 mm wide and 20 mm long, was made, and a pair of plastic hooks which act as pawls were attached to the ends. The strip was suspended from a long plastic ratchet bar, and then immersed in a dilute solution of the surfactant, n -dodecyl pyridinium chloride (C_{12}PyCl) containing $3 \times 10^{-2} \text{ M}$ sodium sulphate (Fig. 1). The gel swelled by a factor of 45 in this solution compared with its dry weight.

We applied a d.c. voltage of 20 V through a pair of planar carbon electrodes (450 mm long, 10 mm wide, 1 mm thick)

Received 26 July; accepted 12 November 1991.

1. *Antennas and Reaction Centers of Photosynthetic Bacteria* (ed. Michel-Beyerle, M. E.) (Springer, Berlin, 1985).
2. *Photoinduced Electron Transfer* (eds Fox, M. A. & Chanan, M.) (Elsevier, Amsterdam, 1988).
3. Grätzel, M. *Heterogeneous Photochemical Electron Transfer* (CRC, Boca Raton, 1989).
4. *Photochemical Energy Conversion* (eds Norris, J. R. & Meisel, D.).
5. Brinker, C. J. & Scherer, G. *Sol-gel Science* (Academic, San Diego, 1990).
6. Slama-Schwok, A., Avnir, D. & Ottolenghi, M. (1991). *J. Am. chem. Soc.* **113**, 3984–3985 (1991).
7. Matsuo, T. *J. Photochem.* **29**, 41–54 (1985).
8. Adar, E., Degani, Y., Goren, Z. & Willner, I. *J. Am. chem. Soc.* **108**, 4696–4700 (1986).
9. Rabani, J. & Sassoon, R. E. *J. Photochem.* **29**, 726 (1985).
10. Margerum, L. D., Murray, R. W. & Meyer, T. J. *J. phys. Chem.* **90**, 728–730 (1986).
11. Moser, J., Grätzel, M. & Gallay, R. *Helv. chim. Acta* **70**, 1596–1604 (1987).
12. Anderson, R. F. *Ber. Bunseng. Phys. Chem.* **80**, 969–972 (1976).
13. Avnir, D., Ottolenghi, M. & Braun, S. in *Supramolecular Architecture in Two and Three Dimensions*, ACS Symp. Series (ed. Bein, T.) (American Chemical Society, in the press).
14. Slama-Schwok, A., Avnir, D. & Ottolenghi, M. *J. phys. Chem.* **93**, 7544–7547 (1989).
15. Slama-Schwok, A., Avnir, D. & Ottolenghi, M. *Photochem. Photobiol.* **54**, 525 (1991).
16. Grellman, K. H., Watkins, A. R. & Weller, A. *J. Luminesc.* **1/2**, 678–692 (1970).

placed in the solution 4 mm above and 16 mm below the top of the ratchet bar, thus producing a field of 10 V cm^{-1} and a current of 15 mA cm^{-2} , and altered the polarity at 2-s intervals. This made the gel move forward with a 'looping' action by repeated bending and stretching. Figure 1 shows successive profiles of the 'gel-looper' as it walks with a constant velocity of 25 cm min^{-1} in the solution.

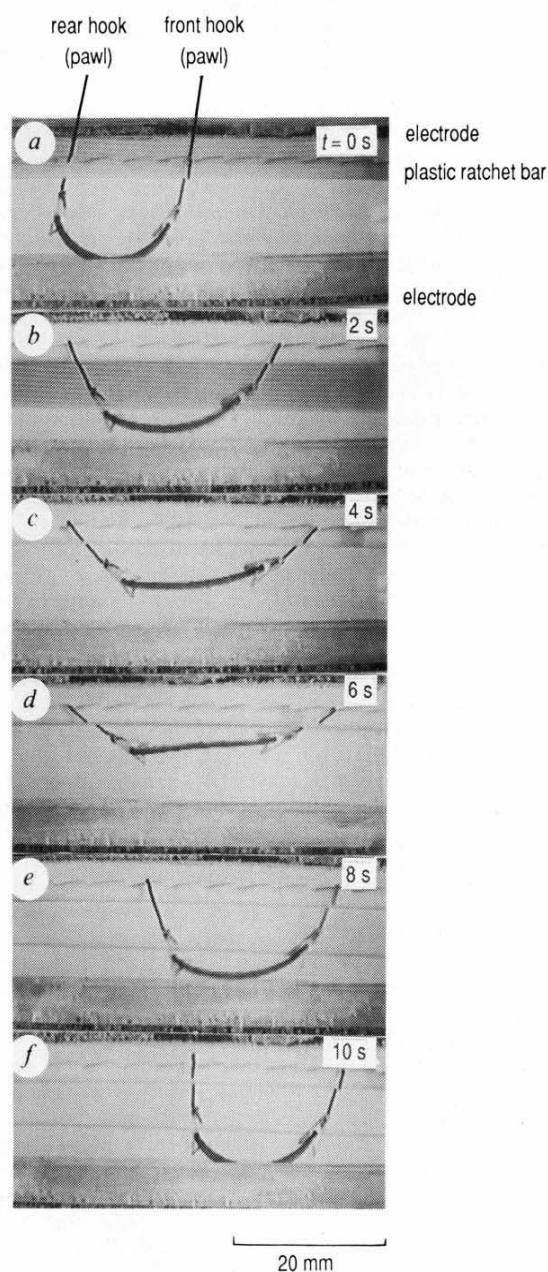


FIG. 1 Time profiles of the polymer gel in action. *a-d*, 'Stretching' process.

FIG. 1 Time profiles of the polymer gel in action. *a-d*, 'Stretching' process. The front hook attached to the gel can slide forward along the plastic ratchet bar, but the rear hook is prevented by the teeth of the ratchet from sliding backwards. *e, f*, Bending process. When the polarity of the electric field is reversed and the gel bends, the rear hook can move forward along the ratchet bar, but the front hook is prevented from sliding backwards. This action being repeated, the gel walks forward. The gel is poly(2-acrylamido-2-methyl propane) sulphonic acid (PAMPS); degree of crosslinking 5 mol%; degree of swelling in the solution, a factor of 45. Electric field, 20 V; current 15 mA cm^{-2} ; electrode distance, 20 mm. Solution: $1 \times 10^{-2} \text{ M}$ *n*-dodecylpyridinium chloride, containing $3 \times 10^{-2} \text{ M}$ sodium sulphate. Total ionic strength: 0.1.

The walking velocity is a function of the applied current, the salt concentrations of surfactant and sodium sulphate, and the molecular size of alkyl chains of surfactant molecules. The velocity increased with increasing current, but there existed optimum sodium sulphate and surfactant concentrations: the maximum velocity of 25 cm min^{-1} occurred at concentrations of $3 \times 10^{-2} \text{ M}$ sodium sulphate and $1 \times 10^{-2} \text{ M}$ C_{12}PyCl at 20 V d.c. Any increase or decrease of concentrations decreased the motility.

The principle of the gel's motility lies in a reversible and cooperative complexing of surfactant molecules on the polymer gel in an electric field, causing the gel to shrink. As is well established⁶, swelling of the ionic gel is characterized by the difference between the osmotic pressure of freely mobile ions in the gel and in the surrounding solution, and their distribution is well described by Donnan equilibrium. In other words, the swelling of the gel is adjusted by varying the number of fixed charges in the gel. Surfactant-hydrogel complexing in one system of this type occurs even in the absence of an electric field, and when the gel is immersed in the surfactant solution, slow but homogeneous shrinkage of the gel is observed. The shrinkage

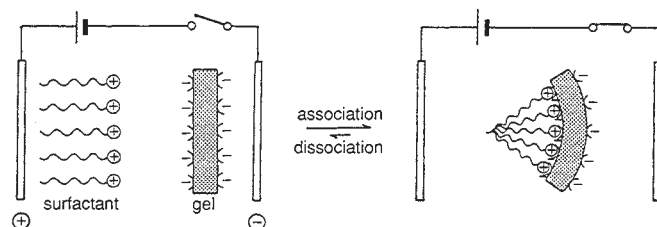


FIG. 2 Schematic illustration of bending mechanism by anisotropic association of surfactant molecules under electric field.

is attributed to the neutralization of negative charges of sulphonates in the gel by forming complexes with surfactant cations decreasing the osmotic pressure difference between the inside of the gel and the surrounding solution. Here, hydrophobic interactions between neighbouring long alkyl chains stabilize the complex and give effective contraction of the gel. We have found that an increase in alkyl chain length sharply increases the association constants of the complex.

The electric field drives and controls the direction of this equilibrium to give anisotropic complex formation. When the d.c. voltage is turned on, the positively charged surfactant molecules move by electrophoresis towards the cathode and form a complex with the negatively charged gel, preferentially on the side of the PAMPS strip facing the anode. This causes anisotropic contraction, bending the gel towards the anode (Fig. 2). When the polarity of the electric field is changed, the surfactant molecules adsorbed on the gel are released and electrophoretically travel towards the anode. Instead, new surfactant molecules form the complex preferentially on the opposite side and straighten the gel. The polymer gel can be made to bend and stretch repeatedly. It thus inches forward along the ratchet bar by means of the hooks from which it is suspended (see Fig. 1).

A moving device of this kind could serve as a new type of 'soft-actuator' or 'molecular machine'. Unlike in motors and hydrodynamic pumps, the motility of this system is produced by the chemical free energy of adsorption of surfactant molecules at the polymer network, with the electrical energy being used to drive the direction and control the state of the equilibrium. Because of the crosslinking bridges in the gel changes in molecular conformation can accumulate and lead to macroscopic shape changes. The gel-actuator also differs from piezoelectric and shape-memory materials where the shape change is caused by a phase transition. The chemomechanical gel has a gentle and flexible action resembling that of muscle rather than motion of metallic mechanical systems. □

Received 29 July; accepted 7 November 1991.

1. Katchalsky, A., Hargitay, B. & Kuhn, W. *Nature* **165**, 514–516 (1950).
2. Katchalsky, A., Oplatka, A. & Steinberg, I. Z. *Nature* **210**, 568–571 (1966).
3. Sussman, M. V. *Nature* **256**, 195–198 (1975).
4. Osada, Y. & Kishi, R. *J. chem. Soc. Faraday Trans.* **85**, 655–662 (1989).
5. Osada, Y., Yamauchi, A. & Umezawa, K. *Bull. chem. Soc. Jpn* **62**, 3232–3238 (1989).
6. Flory, P.J. *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York, 1953).

ACKNOWLEDGEMENTS. We thank T. Yamamoto and Y. Kondo for providing carbon electrodes. This work was supported in part by the Science and Technology Agency, Japan.

Will greenhouse warming lead to Northern Hemisphere ice-sheet growth?

Gifford H. Miller* & Anne de Vernal†

* Center for Geochronological Research, INSTAAR, University of Colorado, Boulder, Colorado 80309-0450, USA

† GEOTOP, Université du Québec à Montréal, CP 8888, Succursale "A", Montréal, Québec, Canada, H3C 3P8

ALTHOUGH model simulations predict a higher mean global temperature by the middle of the next century in response to increased atmospheric concentrations of greenhouse gases¹, the response of the cryosphere to specific changes in latitudinal and seasonal temperature distribution is poorly constrained by modelling^{2,3} or through instrumental measurements of recent variations in snow cover⁴ and ice thickness^{5,6}. Here we examine the recent geological record (130 kyr to present) to obtain an independent assessment of ice-sheet response to climate change. The age and distribution of glacial sediments, coupled with marine and terrestrial proxy records of climate, support arguments that initial ice-sheet growth at the beginning of the last glacial cycle occurred at high northern latitudes (65–80° N) under climate conditions rather similar to present. In particular, the conditions most favourable for glacier inception are warm high-latitude oceans, low terrestrial summer temperature and elevated winter temperature. We find that the geological data support the idea that greenhouse warming, which is expected to be most pronounced in the Arctic and in the winter months, coupled with decreasing summer insolation⁷ may lead to more snow deposition than melting at high northern latitudes⁸ and thus to ice-sheet growth.

Permanent ice bodies influence the heat budget of the atmosphere primarily through albedo feedback and reduction of ocean–atmosphere heat transfer⁹. The volume of land ice also controls sea level, and large continental ice sheets influence the general circulation of the atmosphere. Changes in the area and, to a lesser degree, thickness of land ice can occur rapidly. At the beginning of the last glacial cycle, ice accumulation rates were $\sim 3 \times 10^3 \text{ km}^2 \text{ yr}^{-1}$, accompanied by a corresponding fall in sea level of $\sim 7 \text{ mm yr}^{-1}$ (Fig. 1).

With the recognition of an anthropogenic increase in atmospheric CO_2 , attempt have been made to predict the climate response to these increases and possible changes in the volume of ice on the continents. The question of whether the increased melting of snow due to global warming will be offset by increased snowfall has been raised several times since the late 1970s⁸ and discussed in terms of both model results¹⁰ and direct observations^{5,6}, neither of which has provided a definitive answer^{5,11,12}. Here we address this question from the perspective of the geological record, (1) by identifying the conditions and locus of initial ice-sheet growth in the Northern Hemisphere at the end of the last interglaciation $\sim 120 \text{ kyr}$ ago, and (2) from the behaviour of glaciers and ice caps during the present interglaciation (the past 10 kyr).

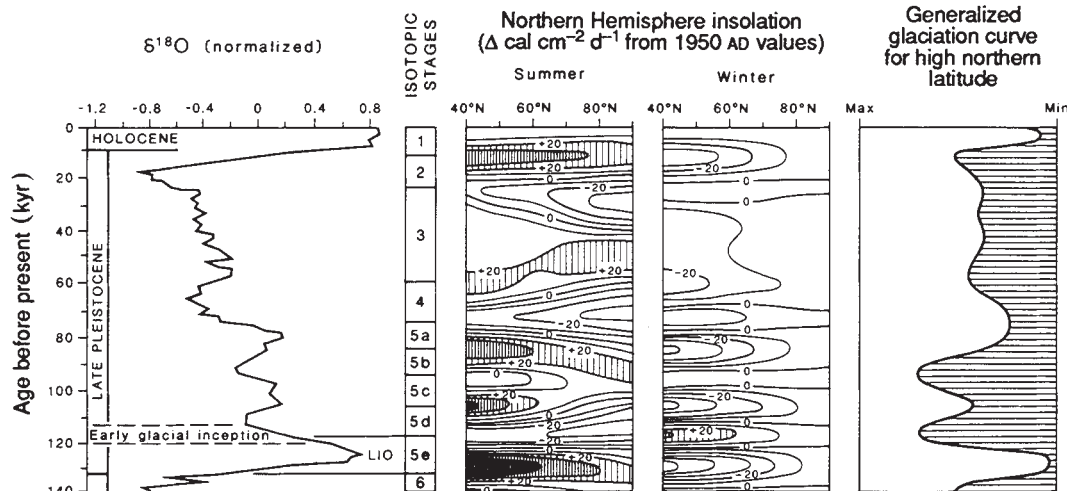
Terrestrial and marine records that cover the inception phase of the last interglacial (~ 120 to 115 kyr BP) strongly suggest that the initial build up of continental ice occurred while terres-

trial climate and sea surface temperatures were close to their interglacial optima. The growth and decay of Quaternary ice sheets can be monitored on a global scale by the $\delta^{18}\text{O}$ stratigraphy in calcareous foraminifera from deep-sea sediments¹³ (Fig. 1). Parallel palaeoclimatic reconstructions have been established from terrestrial and marine palaeoecological indicators that are primarily dependent on temperature. Some sites contain indications of both ice-volume variations and regional palaeoclimates. Marine microfaunal and microfloral assemblages in deep-sea cores from the North Atlantic and adjacent subpolar seas reveal the persistence of optimal sea surface temperatures after the initial increase in $\delta^{18}\text{O}$ that indicates ice accumulation on the continents at the isotope substage 5e/5d boundary^{14–16}. Sites that record the evolution of terrestrial climate and sea level change through the last interglacial are reported from near-shore and terrestrial environments in the United Kingdom¹⁷, the Netherlands¹⁸, Germany¹⁹ and western Norway²⁰, and in a marine core 100 km off the Iberian coast²¹. They show that the marine regression (indicating ice-sheet growth) began while the terrestrial climate remained close to its interglacial optimum. At the onset of the regression (rise in $\delta^{18}\text{O}$), air and sea surface temperatures around the North Atlantic and adjacent subpolar seas were nearly as warm as at present during summer and warmer than present during winter, as shown by quantitative temperature estimates based on marine microfaunal assemblages¹⁵, terrestrial palaeovegetation records²², and isotope composition of the Devon Island and Greenland ice cores^{23,24}.

The location of ice accumulation centres early in the last glaciation can be derived from the record of continental glaciation. Interstratified glacial and near-shore marine sediment along eastern Baffin Island, Canada and northern Greenland indicate that the maximum advance of the last glacial cycle occurred early, during isotope stage 5, while surface water temperatures in Baffin Bay and Labrador Sea were similar to present²⁵. A similar pattern of maximum advance early in the last glaciation is recognized in Svalbard²⁶ and northern Alaska²⁷. Early ice sheet growth over high latitude (65° to 80° N) regions of the Northern Hemisphere is supported by the pattern of ice-rafted detritus (IRD) and reworked microfossils in deep-sea sediments. A relatively high rate of IRD accumulation is recorded during isotope stage 5 in the northwestern North Atlantic²⁸, and reworked pre-Quaternary palynomorphs derived from glacial erosion of the Canadian High Arctic occur with maximum abundance in isotope stage 5 sediments from Baffin Bay and the Labrador Sea¹⁶. In addition, the isotope composition, pollen content and physical characteristics of basal ice in the Canadian Arctic ice cores suggest that the late Pleistocene ice growth started when temperatures were higher than at present, probably during the last interglacial optimum^{23,29}.

The possibility of high-latitude ice-sheet growth during global interglacials is illustrated by well dated records of glacial activity during the Holocene. Nearly the entire northeastern margin of the Laurentide Ice Sheet readvanced or attained its Late Wisconsin maximum during the Cockburn episode 8–9 kyr BP³⁰, when much of the rest of the world was close to the Holocene thermal maximum³¹ and mid-latitude glaciers had already disappeared. Indeed, some local glaciers in Arctic Canada attained their maximum extension of the past 50 kyr during the little Ice Age (~ 1600 – 1900 AD)³². Presumably the climate of the last glacial maximum in the High Arctic was cold and very dry, inhibiting local glacier growth, whereas the warmer but wetter Holocene climate led to ice-sheet growth. Precipitation control of glacier mass balance is well illustrated by cores through Severnaya Zemlya (Soviet Union; mean annual temperature -16°C) ice caps which reveal maximum accumulation rates between 7 and 3 kyr BP³³, that is, during the interval of maximum northward advection of warm oceanic surface waters and moist air masses³¹. In subarctic climate zones at the same latitude (for example, Svalbard; mean annual temperature -5°C), ice cover was markedly reduced during this interval³⁴. Recent lowering

FIG. 1 Changes in global ice volume, insolation and glaciation at high northern latitudes during the past 140,000 years. *a*, The tuned open-ocean $\delta^{18}\text{O}$ record from benthic foraminifera, first-order proxy data for global ice volume changes³⁷; *b*, the seasonal changes in insolation at the top of the atmosphere between 40° and 90°N latitude⁷; and *c*, a generalized glaciation curve derived from regional curves for northeast Canada³⁸, northwest Greenland³⁹ and Svalbard (E. Larsen *et al.*, unpublished data). The early glacial inception at the isotope substages 5e/5d transition spanning less than 10^4 years was marked by ~ 70 m of eustatic fall in sea level¹³ which corresponds



to ice accumulation and rates of fall in sea level of $\sim 3 \times 10^3 \text{ km}^3 \text{ yr}^{-1}$ and $\sim 7 \text{ mm yr}^{-1}$ respectively.

of regional snowlines over Baffin Island, Arctic Canada and Alaska was accompanied by an increase in mean annual temperature, associated with warmer, wetter winters and cooler or unchanged summers^{35,36}. Those regions in which glaciers underwent significant early to middle Holocene expansions have in common a cold and very dry climate (Fig. 2). In such regions, glacier response seems more sensitive to precipitation changes than to changes in air temperature. This assessment is supported by the nonlinear relationship between winter snowfall and the mean summer temperature at the equilibrium line (Fig. 3), which shows that for any incremental increase in snowfall, the concomitant increase in summer temperature required to melt snow is greater in colder regions. Precipitation is no doubt a determinant control in the timing and location of continental glaciations over the high latitudes (cold, dry), where initial ice-sheet

growth occurred during the inception phase of the last glacial cycle.

Predictions of latitudinal and seasonal climate changes based on double the pre-industrial CO_2 concentration in the atmosphere have been made with several general circulation models. Despite differences between the models, all predict that the greatest warming will occur in the polar regions. Most models predict that the circumpolar regions north of 65°N will experience negligible (2°C or less) temperature increases in summer (June, July, August), and a rise of 8–14°C in winter (November–February). Zonally averaged precipitation increases of 20–40% are predicted for the Arctic¹. In addition,

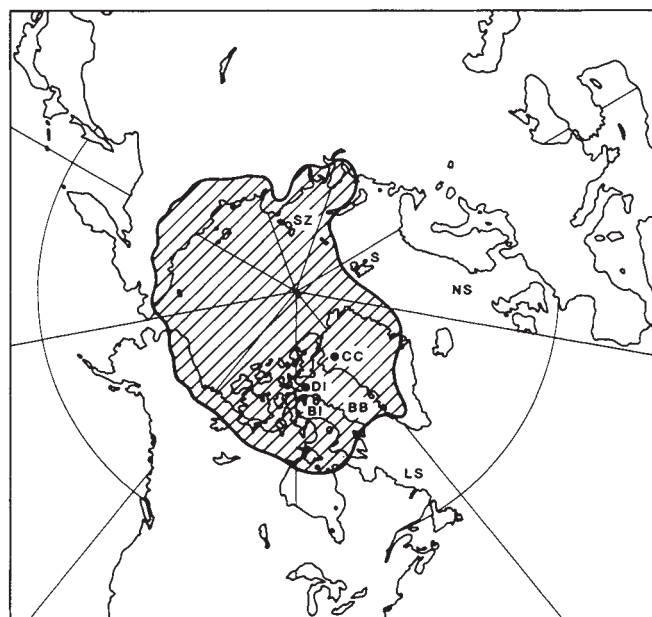


FIG. 2 Polar projection of the Northern Hemisphere. Hatched region encompasses the regions currently cold (mean annual temperature $< -7^\circ\text{C}$) and dry (average annual precipitation < 30 cm water equivalent), which are most sensitive to increased precipitation. Place name abbreviations are as follows: NS, Norwegian Sea; S, Svalbard; BB, Baffin Bay; BI, Baffin Island; LS, Labrador Sea; SZ, Svernavia Zemlya; CC, Camp Century; DI, Devon Island.

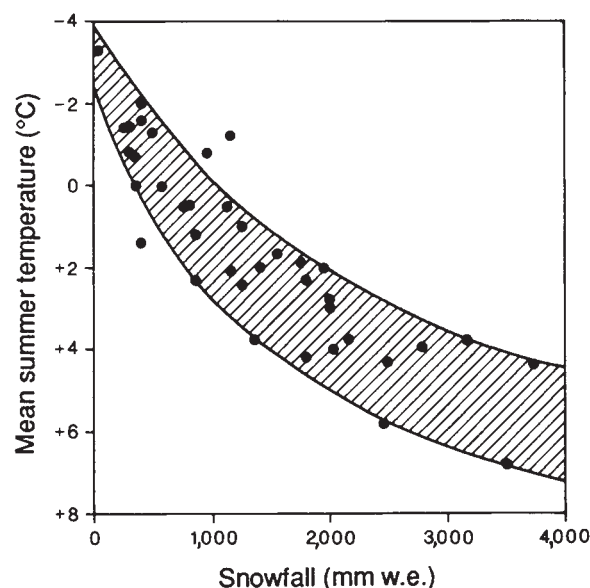


FIG. 3 Snow accumulation (in mm wet equivalent) against mean summer temperature (June, July, August for most sites) at the equilibrium line (that is, the amount of snow melted for a given summer temperature), derived from data summarized in refs 40–42. The relationship corresponds to zero yearly growth of ice. It is defined by an envelope (shaded region) rather than a line because of regional differences in summer insolation, air turbulence and albedo which influence the rate of snow melt in addition to air temperature. The envelope is nonlinear, with a steeper slope in the upper part of the graph that corresponds to cold sites: it demonstrates that for a given increase in snowfall, the concomitant increase of summer temperature required to melt that snow is much greater in the cold regions than at warmer sites.

summer solar radiation at the top of the atmosphere has been decreasing since its interglacial maximum ~9 kyr ago⁷. In the key region between 65° and 70° N, summer radiation is already down to the level corresponding to the inception phase of the last glaciation (Fig. 1), more than half way back to its glacial minimum.

The geological record shows that conditions most conducive for initial ice sheet growth are (1) 'warm' oceanic surface waters during winter, (2) strong meridional atmospheric-oceanic circu-

lation, and (3) low summer air temperature. The predicted change in the global ocean-atmosphere climate system caused by greenhouse warming, coupled with the decreasing incident solar radiation during summer at high northern latitudes, pushes the Earth closer to the conditions that led to the last episode of continental glaciation. Consequently, the possibility that increased concentrations of greenhouse gases in the atmosphere might lead to ice-sheet growth and a concomitant sea level fall (of up to 7 mm yr⁻¹) cannot be lightly dismissed. □

Received 28 May; accepted 25 October 1991.

- Schlesinger, M. E. in *Climate and Geo-sciences* (eds Berger, A., Schneider, S. & Duplessy, J. Cl.) 375-415 (Kluwer, Dordrecht, 1989).
- Saltzman, B. J. *atmos. Sci.* **41**, 2263-2266 (1984).
- Rind, D. *Ann. Glaciol.* **14**, 356 (1990).
- Barry, R. G. *Geol. J.* **20**, 121-127 (1990).
- Zwally, H. J., Brenner, A. C., Major, R. A., Bindshadler, R. A. & Marsh, J. G. *Science* **246**, 1587-1591 (1989); **248**, 288-289 (1990).
- Bentley, C. R. *Eos* **70**, 1002 (1989).
- Berger, A. L. *Quat. Res.* **9**, 137-167 (1978).
- Andrews, J. T. in *Global Changes of the Past* (ed. Bradley, R.) 321-339 (UCAR Office for Interdisciplinary Earth Studies, Boulder, 1991).
- Budyko, M. I. *Tellus* **21**, 611-619 (1969).
- Oerlemans, J. & van der Veen, C. J. *Ice Sheets and Climate* (Reidel, Boston, 1984).
- Untersteiner, N. in *The Global Climate* (ed. Houghton, J. T.) Ch. 8 (Cambridge University Press, 1984).
- Douglas, B. C. et al. *Science* **248**, 288 (1990).
- Shackleton, N. J. *Quat. Sci. Rev.* **6**, 183-190 (1987).
- Ruddiman, W. F. & McIntyre, A. *Science* **204**, 173-175 (1979); **212**, 617-627 (1981).
- CLIMAP Project members *Quat. Res.* **21**, 123-224 (1984).
- Hillaire-Marcel, C. & de Vernal, A. *Géogr. phys. Quat.* **43**, 263-290 (1989).
- Sparks, B. W. & West, R. G. *Phil. Trans. R. Soc. Lond.* **B258**, 1-30 (1970).
- Zagwijn, W. H. *Geol. Mijnbouw* **62**, 437-450 (1983).
- Knudsen, K. L. *Geol. J.* **A86**, 27-47 (1985).
- Mangerud, J., Sønstegeard, E., Sejrup, H. P. & Haldorsen, S. *Boreas* **10**, 137-208 (1981).
- Turon, J.-L. *Nature* **309**, 673-676 (1984).
- Guiot, J. *Palaeogeogr. Palaeoclim. Palaeoecol.* **80**, 49-69 (1990).
- Koerner, R. M. *Science* **244**, 964-968 (1989).

- Reeh, N. in *Quaternary Geology of Canada and Greenland* (ed. Fulton, R. J.) Ch. 14 (Geological Survey Canada, Ottawa, 1989).
- Miller, G. H., Funder, S., de Vernal, A. & Andrews, J. T. in *Inception of the Last Glaciation* (eds Clark, P. U. & Lea, P.) (Geological Society of America Special Paper, in the press).
- Miller, G. H., Sejrup, H. P., Lehman, S. J. & Forman, S. L. *Boreas* **18**, 273-296 (1989).
- Hamilton, T. D. in *Glaciation in Alaska: The Geologic Record* (eds Hamilton, T. D., Reed, K. M. & Thorson, R. M.) 9-50 (Alaska Geological Survey, Anchorage, 1986).
- Ruddiman, W. F. *Geol. Soc. Am. Bull.* **88**, 1813-1827 (1977).
- Koerner, R. M., Bourgeois, J. C. & Fisher, D. A. *Ann. Glaciol.* **10**, 85-91 (1988).
- Andrews, J. T. & Ives, J. D. *Arct. Alpine Res.* **10**, 617-633 (1978).
- COHMAP members *Science* **241**, 1043-1052 (1988).
- Miller, G. H. *Geology* **4**, 502-504 (1976).
- Potapenko, V. Yu., Klementyev, O. L. & Nikolayev, V. I. *Int. Symp. Glaciers-Ocean-Atmosphere Interactions* (Leningrad, 1990).
- Svendsen, J. I. & Mangerud, J. *Clim. Dynam.* (in the press).
- Bradley, R. S. & Miller, G. H. *Nature* **237**, 385-387 (1972).
- Mayo, L. R. & March, R. S. *Ann. Glaciol.* **14**, 191-194 (1990).
- Martinson, D. G. et al. *Quat. Res.* **27**, 1-29 (1987).
- Andrews, J. T. *Phil. Trans. R. Soc. Lond.* **B318**, 645-660 (1988).
- NORDQUA 86 Participants *Géogr. phys. Quat.* **43**, 1740-1750 (1989).
- Loewe, F. *Arct. Alp. Res.* **3**, 331-344 (1971).
- Ohmura, A. *Climate and Energy Balance on Arctic Tundra* (Zürcher Geographische Schriften, Zurich, 1981).
- Mangerud, J. in *Abrupt Climate Change* (eds Berger, W. H. & Labeyrie, L. D.) 163-171 (Reidel, Boston, 1987).

ACKNOWLEDGEMENTS. This work was stimulated by a NATO Advanced Studies Workshop. Our research has been sponsored by the US NSF (G.H.M.) and NSERC-Canada and Fonds FCAR-Quebec (A.de V.).

Drowned islands downstream from the Galapagos hotspot imply extended speciation times

D. M. Christie*, R. A. Duncan*, A. R. McBirney†, M. A. Richards‡, W. M. White§, K. S. Harpp§, & C. G. Fox||

* Oregon State University, College of Oceanography, Corvallis, Oregon 97331-5503, USA

† University of Oregon, Department of Geology, Eugene, Oregon 97403, USA

‡ University of California at Berkeley, Department of Geology and Geophysics, Berkeley, California 94720, USA

§ Cornell University, Department of Geology, Ithaca, New York 14853, USA

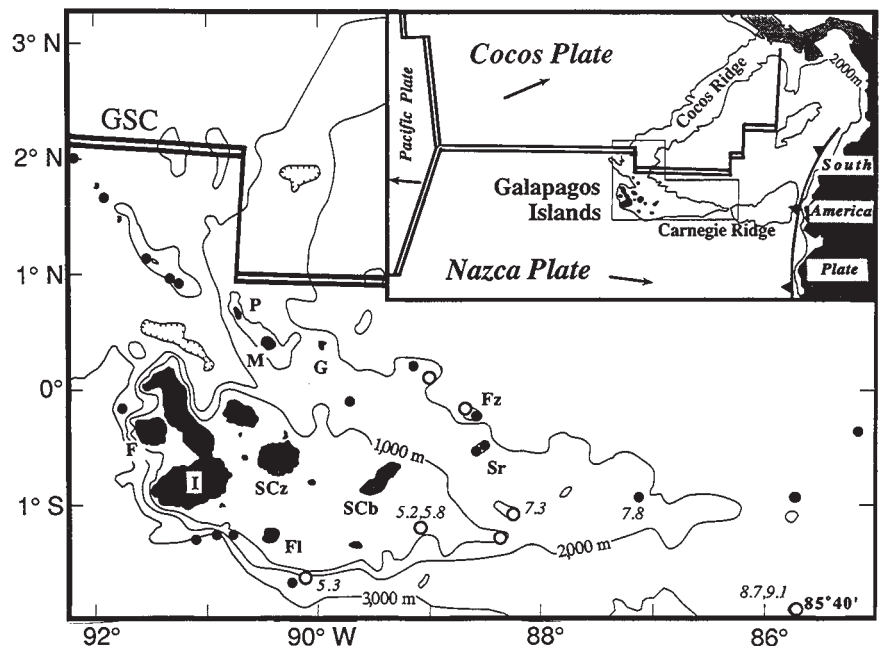
|| NOAA, Hatfield Marine Science Center, Newport, Oregon 97365, USA

THE volcanic islands of the Galapagos archipelago are the most recent products of a long-lived mantle hotspot^{1,2}. Little is known, however, of the submarine Galapagos platform on which the islands are built, or of the Cocos and Carnegie submarine ridges produced by past motion of the Cocos and Nazca plates across the hotspot^{3,4}. In 1990 we surveyed selected areas around the Galapagos platform and as far east as 85° 30' W on the Carnegie ridge, where we dredged abundant well-rounded basalt cobbles from a small seamount with a terraced summit region. Cobbles were also dredged from several other seamounts. We interpret these features, especially the presence of cobbles, as evidence for erosion near sea level and conclude that these seamounts were volcanic islands before subsiding to their present depths. Radiometric ages for these drowned islands range from 5 to 9 Myr, consistent with predicted plate motions. They indicate that the time available for speciation of Galapagos organisms is much longer than the age range of the existing islands.

The oldest known and best documented of the drowned islands, referred to here as the 85° 40' W seamount, seems to have formed over the hotspot at least 9 Myr ago. This small seamount is the middle of three similar east-west aligned edifices, previously surveyed along 2° S latitude on the southern flank of the Carnegie ridge (ref. 5, and P. Lonsdale, unpublished data from the Ceres 04 expedition). The basal diameter of this seamount is ~10 km, roughly the size of the smaller present-day islands such as Isla Pinzón. It rises steeply to a generally flat summit region at 2,000 m depth with a small peak towards its northwest edge (Fig. 2). The summit morphology can be interpreted in terms of a number of wave-cut terraces at different depths and a residual spire. Much stronger evidence for wave action is provided by the recovery of abundant, highly rounded basalt pebbles and cobbles in a dredge haul from the upper north slope. All the cobbles recovered are composed of aphyric basalt with greenish-brown discoloration due to low-temperature seawater alteration. Many have large, irregularly shaped, interconnected vesicles of a type that is common in subaerial flows but which we have not observed in lavas erupted in deep water. The four analysed samples from this seamount are moderately evolved, low-TiO₂, low-Na₂O tholeiitic basalts closely resembling the 'low-K₂O' basalts of the younger islands^{6,7}. Isotopic ratios (⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd) for these samples are also consistent with their origin at the Galapagos hotspot.

Basalt pebbles and cobbles were also recovered at a number of other sample sites (Table 1 and Fig. 1). A second large, previously unmapped seamount to the north of Sunray seamount has been named 'Fitzroy' after the captain of HMS *Beagle*. It yielded well-rounded cobbles, although it lacks morphologic features consistent with sea-level erosion, perhaps indicating more rapid subsidence. In addition, three sample sites on smaller edifices, which were not sufficiently well surveyed for morphologic studies, yielded rounded pebbles of various sizes. In all cases, chemical compositions of basalts lie within the range

FIG. 1 Generalized bathymetry of the Galapagos area. The Galapagos platform is the broad, generally flat region bounded by the 1,000-m contour. Open circles, dredge sites believed to have sampled drowned islands; closed circles, dredge sites that recovered pillow basalts from seamounts that may or may not have been islands, but yielded no evidence of subaerial processes. Numbers are $^{40}\text{Ar}/^{39}\text{Ar}$ radiometric ages. Island and seamount names are abbreviated: F, Fernandina; FI, Floreana; Fz, Fitzroy seamount; G, Genovesa; I, Isabela; M, Marchena; P, Santa Cruz; SCb, San Cristobal; SCz, Santa Cruz; Sr, Sunray seamount. GSC is the Galapagos spreading centre.



of samples from the Galapagos Islands. These occurrences imply the existence of a number of other drowned islands on the eastern end of the Galapagos platform and western Carnegie ridge.

The production of volcanic ridges and associated seamount chains as lithospheric plates move over stationary volcanic hotspots is a well-documented phenomenon⁸⁻¹⁰. In the case of the Galapagos hotspot, the Cocos and Carnegie ridges record the motions of the Cocos and Nazca plates, respectively, away from a hotspot that was located beneath the Cocos-Nazca spreading centre until ~5 Myr ago³. At that time, the spreading centre moved north, away from the hotspot, severing the connection between Cocos ridge and the hotspot. In the present Galapagos archipelago, volcanoes have been active on almost all the islands during the past 100,000 years, but the maximum measured ages increase systematically eastward from the youngest volcano, Isla Fernandina (Fig. 3). This island age progression is consistent with predicted eastward Nazca plate motion at 40–50 mm yr⁻¹ over the past 3 Myr (refs 11, 12). Earlier than this, the Nazca plate may have moved rather faster (~75 mm yr⁻¹) over the hotspot¹³ (Fig. 3). The dated cobbles from the 85° 40' W seamount are ~2 Myr younger than the maximum age (~11 Myr) predicted for this plate velocity, well within the lifespan of many of the present islands. Particularly when we consider that the cobbles probably represent the younger products of this volcano,

the measured ages clearly support the idea that this seamount was a volcanic island that first became active over the Galapagos hotspot ~10–11 Myr ago. Dated samples from four other Carnegie ridge seamounts (Table 1, Fig. 3) define a clear age-distance trend consistent with prolonged hotspot activity.

Although the existence of drowned islands along Carnegie ridge is a predictable consequence of plate motion over the Galapagos hotspot², the actual documentation of their existence may help to resolve a controversy among evolutionary biologists. Some have argued that the geological youth of the present islands (≤ 3 Myr) requires that all adaptive radiation has occurred within that period¹⁴. Others have maintained, on the basis of molecular studies of proteins, that much longer periods are required for the divergence of some stocks, most notably the iguanas, which may have required as much as 15–20 Myr for divergence from a common gene pool¹⁵.

Our geological observations and radiometric data indicate that there have been islands present over the Galapagos hotspot for at least 9 Myr and probably much longer. The geometry of the Cocos and Carnegie ridges is consistent with hotspot volcanism in the Galapagos region for the past 15–20 Myr, and large portions of the Caribbean plate (of age 80–90 Myr) have been linked to the initiation of Galapagos volcanism through compositional studies¹⁶ and plate reconstructions¹³. We consider it

TABLE 1 Locations of dated samples and others indicative of emergent conditions

Dredge	Latitude	Longitude	Depth (m)	Age/Sample no.*	Notes
1	2° 00.5' S	85° 39.7' W	2,500–2,200	8.7 ± 0.4/PL02-1-12 9.1 ± 0.4/PL02-1-46	Small flat-top terraced seamount. Well-rounded vesicular and massive basalt pebbles and cobbles.
4	0° 55.9' S	87° 07.0' W	1,800–1,300	7.8 ± 1.2/PL02-4-19	Small seamount. Angular basalt pieces only.
5	1° 17.3' S	88° 21.4' W	875–575		Small seamount. Cobbles of vesicular basalt.
8	1° 07.0' S	88° 15.8' W	1,000–900	7.3 ± 1.5/PL02-8-1	Small steep seamount. Two cobbles vesicular basalt.
11	0° 12.0' S	88° 40.5' W	1,300–700		Large seamount (Fitzroy). Cobbles vesicular and massive basalt.
13	0° 05.8' N	89° 02.9' W	1,400–700		Composite seamount. Two orange weathered finely vesicular basalt cobbles in addition to pillow basalt fragments.
17	01° 11.4' S	89° 06.6' W	650–300	5.8 ± 0.8/PL02-17-4 5.3 ± 0.4/PL02-17-1	East-west ridge. Well-rounded basalt pebbles. Branching coral.
20	1° 37.7' S	90° 10.7' W	1,050–300		Small seamount. Basalt pebbles (1.3 cm).

* $^{40}\text{Ar}/^{39}\text{Ar}$ plateau ages.

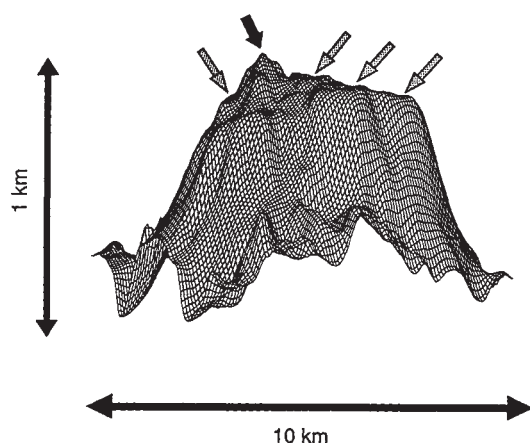


FIG. 2 Perspective view of the 85° 40' W seamount. Features of the summit region that we interpret as consistent with its origin as a drowned island are wave-cut terraces (stippled arrows) and a residual spire (solid arrow). Based on gridded Seabeam bathymetric data. View angle is roughly horizontal from the southwest.

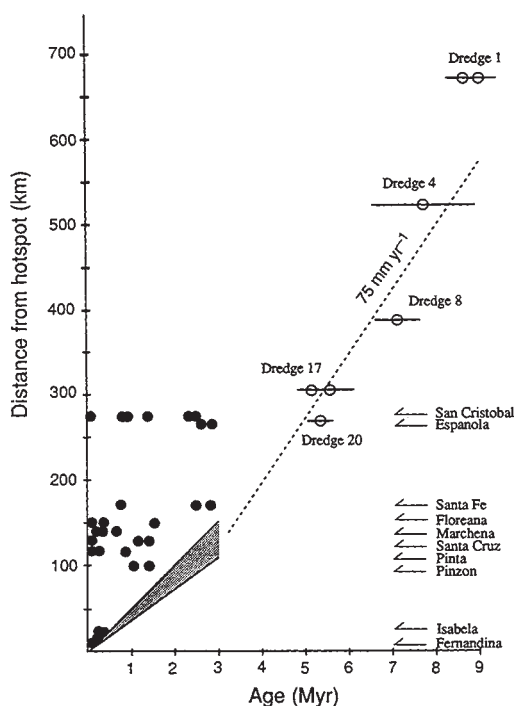


FIG. 3 Radiometric ages against distance from the centre of the Galapagos hotspot, beneath Isla Fernandina. Filled circles are K/Ar ages from island samples, open circles are $^{40}\text{Ar}/^{39}\text{Ar}$ ages from dredged seamount samples (1σ error bars). Reported ages (Table 1) are weighted means of concordant ages from gas fractions released at sequential heating steps. Predicted velocities for the Nazca plate over the hotspot^{11,12} fall within the shaded region for the past 3 Myr, whereas a faster motion (dotted line) is required for earlier times¹³.

likely that islands have existed through the entire 80–90 Myr history of hotspot activity. □

Received 8 July; accepted 6 November 1991.

1. Morgan, W. J. *Nature* **230**, 42–43 (1971).
2. Morgan, W. J. *Am. Assoc. Petrol. Geol. Bull.* **56**, 203–213 (1972).
3. Hey, R. N. *Geol. Soc. Am. Bull.* **88**, 1404–1420 (1977).
4. Cox, A. in *Patterns of Evolution in Galapagos Organisms* (eds Bowman, R. I. & Leviton, A. E.) 11–23 (American Association for the Advancement of Science, Washington DC, 1983).
5. Malfait, B. T. thesis, Oregon State Univ. (1975).
6. Geist, D. J., McBirney, A. R. & Duncan, R. A. *Geol. Soc. Am. Bull.* **97**, 555–566 (1986).

7. Baitis, H. W. & Swanson, F. J. *Nature* **259**, 195–197 (1976).
8. Dalrymple, G. B., Lanphere, M. A. & Clague, D. A. in *Init. Rep. Deep-Sea Drilling Project 55* (eds Jackson, E. D. et al.) 659–676 (US Government Printing Office, Washington, 1981).
9. O'Connor, J. M. & Duncan, R. A. *J. geophys. Res.* **95**, 17475–17502 (1990).
10. Duncan, R. A. & Clague, D. A. in *The Ocean Basins and Margins 7A* (eds Nairn, A. E. M., Stehli, F. G. & Uyeda, S.) 89–121 (Plenum, New York, 1985).
11. Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331–5354 (1978).
12. Gripp, A. E. & Gordon, R. G. *Geophys. Res. Lett.* **17**, 1109–1112 (1990).
13. Duncan, R. A. & Hargraves, R. B. in *The Caribbean–South American Plate Boundary and Regional Tectonics*, *Geol. Soc. Am. Mem.* **162** (eds Bonini, W. E., Hargraves, R. B. & Shagam, R.) 81–94 (1985).
14. Hickman, C. S. & Lipps, J. H. *Science* **227**, 1578–1580 (1985).
15. Wiles, J. S. & Sarich, V. M. in *Patterns of Evolution in Galapagos Organisms* (eds Bowman, R. I. & Leviton, A. E.) 177–186 (American Association for the Advancement of Science, Washington DC, 1983).
16. Sen, G., Hickey-Vargas, R., Waggoner, D. G. & Maurasse, F. *Earth planet. Sci. Lett.* **87**, 423–437 (1988).

ACKNOWLEDGEMENTS. We thank the government of Ecuador, the Parque Nacional Galapagos and the Charles Darwin Research Station for permission to carry out the field work, the officers and crew of the RV *Thomas Washington* for cooperation and the NSF for funding.

Subtractive and divisive adaptation in the human visual system

Robert J. Snowden & Stephen T. Hammett

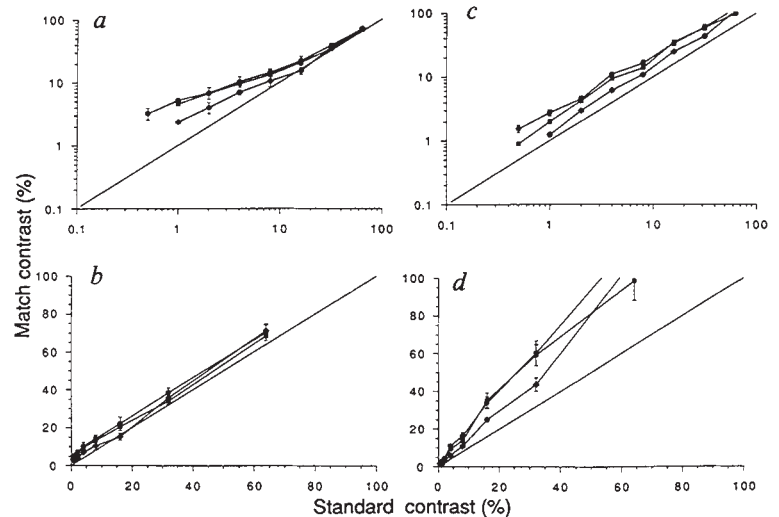
School of Psychology, University of Wales College of Cardiff, PO Box 901, Cardiff CF1 3YG, UK

SENSORY systems can adapt to the conditions imposed on them¹. In the visual system, adapting to a pattern increases the threshold of the ability to see that pattern, and reduces the perceived contrast of the pattern above threshold^{2–4}. Most neurons of the striate cortex reduce their responsiveness after being stimulated for some time by a high-contrast pattern^{5–7}. Such an effect may lie behind these psychophysical adaptation phenomena^{2–4}. These adaptation effects have been reported to be confined to patterns of similar orientation, which is understandable in that the visual neurons that adapt are only excited by a small range of orientations⁸. Neurophysiological evidence suggests that neurons with different orientation preferences have inhibitory interconnections^{9–13}. It is therefore of interest to explore the possible effects of these connections on perception. Here we show that adapting to a horizontal pattern can reduce the perceived contrast of a vertical test pattern more than a horizontal test pattern. These 'cross-orientation' effects are modelled by a division-like process, whereas the more normal 'similar-orientation' effects are modelled by a subtractive process.

We adapted one half of the subject's visual field to a high-contrast sinusoidal grating, and had the subject match the contrast of two gratings. One grating was presented in the unadapted half of the visual field, and the other was presented to the adapted side of the field.

The contrast required to match various standard contrasts are plotted as a function of the standard contrast in Fig. 1. The line of slope 1 represents a perfect match. Results gathered without adaptation (not displayed) fell on this line within experimental error. The results gathered after adaptation are represented by the solid symbols. The results when the adapting and test patterns have the same orientation are plotted in *a* and *b* (*a* and *c* show the results on logarithmic axes, whereas the *b* and *d* present the results on linear axes). These results agree with those recently reported by Georgeson¹⁴ in that the loss of contrast appears most drastic at low contrasts when plotted on logarithmic axes. The loss of contrast appears to be 'subtractive', in that at each standard contrast a set amount of contrast is lost. Thus, on linear axes this line appears to parallel the ideal match line. The results when adapting to a horizontal pattern, but testing with a vertical pattern are illustrated *c* and *d*. It can be seen that more contrast is always required on the adapted side to match the standard pattern. On logarithmic axes the function appears to parallel the ideal match, whereas on linear axes the

FIG. 1 The contrast required to match a standard contrast is plotted after adaptation to patterns of the same orientation (*a, b*) and after adaptation to patterns of orthogonal orientations (*c, d*). *a, c*, Data plotted on logarithmic axes; *b, d* data on linear axes. The different symbols refer to three different subjects (●, R.S.; ◆, J.O.; ■, S.H.). Error bars, ± 1 s.e. The stimuli were all sinusoidal grating patterns of 5 c/d that counterphased at 2 Hz (to avoid after-images). The adapting pattern was presented in a window of 5° diameter centred 3.5° to the right of the fixation point, and had a contrast of 60%. The mean luminance of the stimuli, and of the surrounding screen, was 200 cd m⁻². The two 'test' patterns were of 3° diameter and each was centred 3.5° to the left or right of the fixation point, respectively. After an adaptation period (60 s) the subject was presented with the two test patterns (a standard contrast to the left, and a variable-contrast to the right) for 0.5 s and made a judgment as to which had the most perceived contrast. This judgment was used to adjust the contrast of the next variable-test according to a QUEST procedure. Before each test there was a 'top-up' period of 6 s.



loss of contrast appears most drastic at high contrasts. This loss of contrast appears to be 'divisive' (or 'multiplicative') in that the adapted side always needs a multiple of the standard contrast to achieve a perceptual match.

From the above results it seems that there may be two forms of adaptation: a subtractive process for patterns of the same orientation, and a divisive process for patterns of orthogonal orientations. To answer the question of what is similar and what is different, the predictions of these two forms of adaptation have to be considered.

Figure 2 illustrates graphically the predictions of a model of the two processes. The output of a 'channel' is plotted as a function of the contrast of the stimulating pattern. The solid line represents the output of a visual channel before any adaptation. The channel is unresponsive below a certain contrast, and is then a linear function of contrast over some range before saturating at high contrasts¹⁵⁻¹⁷. The broken line shows the workings of a subtractive process. This manipulation results in a change in threshold, and, at suprathreshold levels, the unadapted and adapted contrast are a constant distance apart. The dotted line illustrates the divisive process. The prediction is quite different in that threshold contrast is unchanged, and the functions become more different at higher contrasts. Hence, the subtractive process is most easily seen at threshold, whereas the divisive process is most easily seen at high-contrast levels. We therefore collected data on the effects of varying the orientation of the adapting pattern in (1) threshold elevation and (2) perceived contrast of a high-contrast grating, after adapting to a high-contrast grating.

Figure 3c shows the threshold elevation (solid symbols) and the loss of perceived contrast (open symbols) as a function of adapting orientation. The two functions have very different tuning characteristics. Threshold elevation is most affected when adapting and test patterns had no orientation difference (as has been noted on many occasions²⁻⁴) and quickly fell off as the orientation difference increased. There was no threshold elevation for orthogonal patterns. The effect of these same adapting patterns on the contrast of suprathreshold gratings has a very different tuning function. There is little tuning, with some hint of a maximal effect when the adapting pattern is around 45° different to the test pattern. The underlying tuning of the 'divisive' effect is hard to ascertain as one would need to know how much of the match elevation is due to the subtractive effect.

Blakemore and colleagues^{3,4} have shown that, after adaptation, both the elevation of threshold and loss of suprathreshold contrast are both tuned to the adapting orientation (that is the effect is largest for parallel adapt and test patterns). Our results seem to contradict this, in that the least loss of suprathreshold contrast was found when the test and adapting patterns are

parallel. We were therefore concerned that our results might be specific to the particular parameters used. This was not the case. Measurements were made at a number of spatial frequencies (1–15 cycles per degree) and at several frequencies (0–8 Hz) (Fig. 3a, b). The 'cross-orientation' adaptation effect was present for all stimuli used. It was also found when the screen was tilted so as to make the adapting grating at 45° and the test at -45°. The contradiction between the two sets of results is puzzling but the studies do have large differences in luminance, size and adapting regimes, all of which may be important.

Interactions between patterns of very different orientation have been previously reported. For instance, adapting and testing with patterns which are near to orthogonal can alter the perceived orientation of the test pattern, the 'indirect effect'¹⁸. Of particular interest is a report by Heeley¹⁹ that the perceived spatial frequency of a pattern can be altered by adapting to an orthogonal grating, without altering the observer's ability to detect the pattern. Our results offer an explanation of this effect. Adapting at the orthogonal orientation will reduce the suprathreshold responsiveness of the channels at the adapting spatial frequency. Hence there will be no effect on threshold, but at suprathreshold levels the contribution of the spatial frequency channels which are the same as the adapting pattern will be reduced causing a shift in perceived spatial frequency away

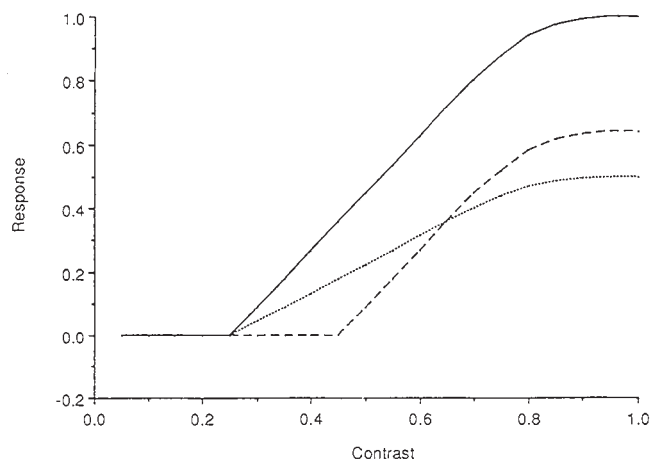


FIG. 2 Idealized contrast versus response functions before (solid line) and after subtractive or divisive adaptation. The broken line shows the workings of a subtractive process. It was produced by shifting the whole function down (subtracting a set amount from the response), and points which would have been negative are set to zero. The dotted line illustrates the divisive process, produced by dividing the response by a set amount.

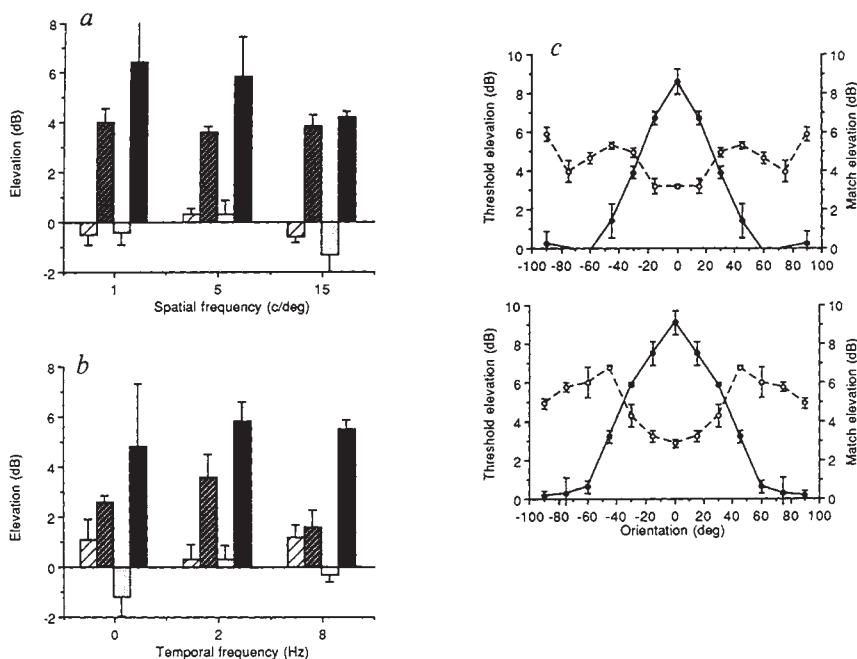


FIG. 3 *a*, The amount of threshold elevation (light bars) and the loss of contrast for a high contrast grating (dark bars) are plotted for three spatial frequencies after adapting to a horizontal pattern and testing with a vertical pattern (all patterns had a temporal frequency of 2 Hz). All other conditions are as in Fig. 1. Striped bars, subject R.S.; speckled bars, subject J.O. Error bars 1 s.e. The results are consistent in showing little or no change in threshold but a loss of suprathreshold contrast. *b*, As in *a* but for a range of temporal frequencies (spatial frequencies = 3 cycles per degree). *c*, The effects of changing the angle between the adapting and test patterns on threshold detection (solid symbols) and on the perceived suprathreshold contrast of a grating of 32% contrast (open symbols). Results for subject J.O. are shown in the upper panel and those for R.S. in the lower panel. Error bars, ± 1 s.e. All conditions are as in Fig. 1, except that the adapting pattern had a contrast of 90%. To measure thresholds the same procedures were used as in Fig. 1, except that the contrast of the pattern on the unadapted side was set to 0% and subjects gave a yes/no decision on the visibility of the test pattern. This decision was used to drive the QUEST procedure.

from the adapting spatial frequency. The tuning of the spatial frequency shift across orientation is different from the classical effect²⁰, but corresponds well with the spatial frequency tuning of the divisive adaptation (R.J.S., manuscript in preparation).

The actual mechanisms of adaptation (in terms of the classic similar-orientation adaptation, and the new cross-orientation adaptation) have yet to be fully resolved, though one popular notion is that it involves inhibition over a protracted time course¹⁵. This idea is attractive as explains the results of the experiments of Greenlee and Magnussen²¹. They show that adapting to two different orientations (either simultaneously or successively) can cause less adaptation than to either one alone. This can be explained if each of the adapting patterns is inhibiting the other. Our results suggest that this will be a divisive process. The current ideas receive support from electrophysiological evidence. For example, Morrone and Burr²² measured visually evoked potentials as a function of contrast in the presence and absence of parallel or orthogonal masking patterns. They show that parallel masks produce a shift in threshold and no change in slope, whereas orthogonal masks tend to change the slope without affecting threshold, thus complementing our results.

Cells of the striate cortex with different orientation preferences may exert inhibitory influences upon one another⁹⁻¹³. Are these inhibitory influences responsible for the present cross-orientation results? At first this might seem unlikely. If these inhibitory influences are 'adapted-out' we might expect the test pattern to appear to have a greater contrast. But if, as suggested above, adaptation reflects prolonged inhibition then the results are understandable. Recent reports of similar divisive inhibition between direction-tuned cells of area MT of the monkey²³ (see also ref. 24 for results in area 17 of the cat) might suggest that this divisive cross-dimensional inhibition is a pervasive influence throughout sensory systems. □

- Morrone, M. C., Burr, D. C. & Maffei, L. *Proc. R. Soc. B* **216**, 335-354 (1982).
- Ramoa, A. S., Shadlen, M., Skottun, B. C. & Freeman, R. D. *Nature* **321**, 237-239 (1986).
- Bonds, A. B. *Vis. Neurosci.* **2**, 41-55 (1989).
- Georgeson, M. A. *Spat. Vision* **1**, 103-112 (1985).
- Dealy, R. S. & Tolhurst, D. J. *J. Physiol., Lond.* **241**, 261-270 (1974).
- Albrecht, D. G. & Hamilton, D. B. *J. Neurophys.* **48**, 217-237 (1982).
- Tolhurst, D. J., Movshon, J. A. & Thompson, I. D. *Exp. Brain Res.* **41**, 414-419 (1981).
- Gibson, J. J. & Radner, M. J. *exp. Psychol.* **20**, 453-467 (1937).
- Heeley, D. W. *Vision Res.* **29**, 1229-1236 (1979).
- Blakemore, C. B. & Sutton, P. *Science* **166**, 245-247 (1971).
- Greenlee, M. W. & Magnussen, S. *Vision Res.* **28**, 1303-1310 (1988).
- Morrone, M. C. & Burr, D. C. *Nature* **321**, 235-237 (1986).
- Snowden, R. J., Treue, S., Erickson, R. E. & Andersen, R. A. *J. Neurosci.* **11**, 2768-2785 (1991).
- Dean, A. F., Hess, R. F. & Tolhurst, D. J. *J. Physiol., Lond.* **308**, 84 (1980).

ACKNOWLEDGEMENTS. We thank J. Orchard for technical assistance.

Tit for tat in heterogeneous populations

Martin A. Nowak* & Karl Sigmund†

* Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

† Institut für Mathematik, Universität Wien, Strudlhofg. 4, A-1090 Wien, Austria

THE 'iterated prisoner's dilemma' is now the orthodox paradigm for the evolution of cooperation among selfish individuals. This viewpoint is strongly supported by Axelrod's computer tournaments, where 'tit for tat' (TFT) finished first¹. This has stimulated interest in the role of reciprocity in biological societies¹⁻⁸. Most theoretical investigations, however, assumed homogeneous populations (the setting for evolutionarily stable strategies^{9,10}) and programs immune to errors. Here we try to come closer to the biological situation by following a program⁶ that takes stochasticities into account and investigates representative samples. We find that a small fraction of TFT players is essential for the emergence of reciprocity in a heterogeneous population, but only paves the way for a more generous strategy. TFT is the pivot, rather than the aim, of an evolution towards cooperation.

The simple 'prisoner's dilemma' is a game with two players, each having two options, *C* (to cooperate) and *D* (to defect). We use Axelrod's payoff values: if both players cooperate, both obtain $R = 3$ points; if both defect, each receives $P = 1$ point; if one player defects and the other cooperates, the defector gets $T = 5$ points and the cooperator $S = 0$. Clearly strategy *D* is

Received 12 August; accepted 25 October 1991.

- Mollon, J. D. *New Scientist* **61**, 479-482 (1974).
- Gilinsky, A. S. *J. opt. Soc. Am.* **58**, 13-17 (1968).
- Blakemore, C. & Nachmias, J. *J. Physiol., Lond.* **213**, 157-174 (1971).
- Blakemore, C., Muncey, J. P. J. & Ridley, R. M. *Vision Res.* **13**, 1915-1931 (1973).
- Maffei, L., Fiorentini, A. & Bisti, S. *Science* **182**, 1036-1038 (1973).
- Movshon, J. A. & Lennie, P. *Nature* **278**, 850-852 (1979).
- Ohzawa, I., Sclar, G. & Freeman, R. D. *Nature* **298**, 266-268 (1982).
- Hubel, D. & Weisel, T. *J. Physiol., Lond.* **195**, 215-243 (1968).
- Benevento, L. A., Creutzfeldt, O. D. & Kuhnt, U. *Nature* **238**, 124-126 (1972).
- Sillito, A. M. *Trends Neurosci.* **2**, 196-198 (1979).

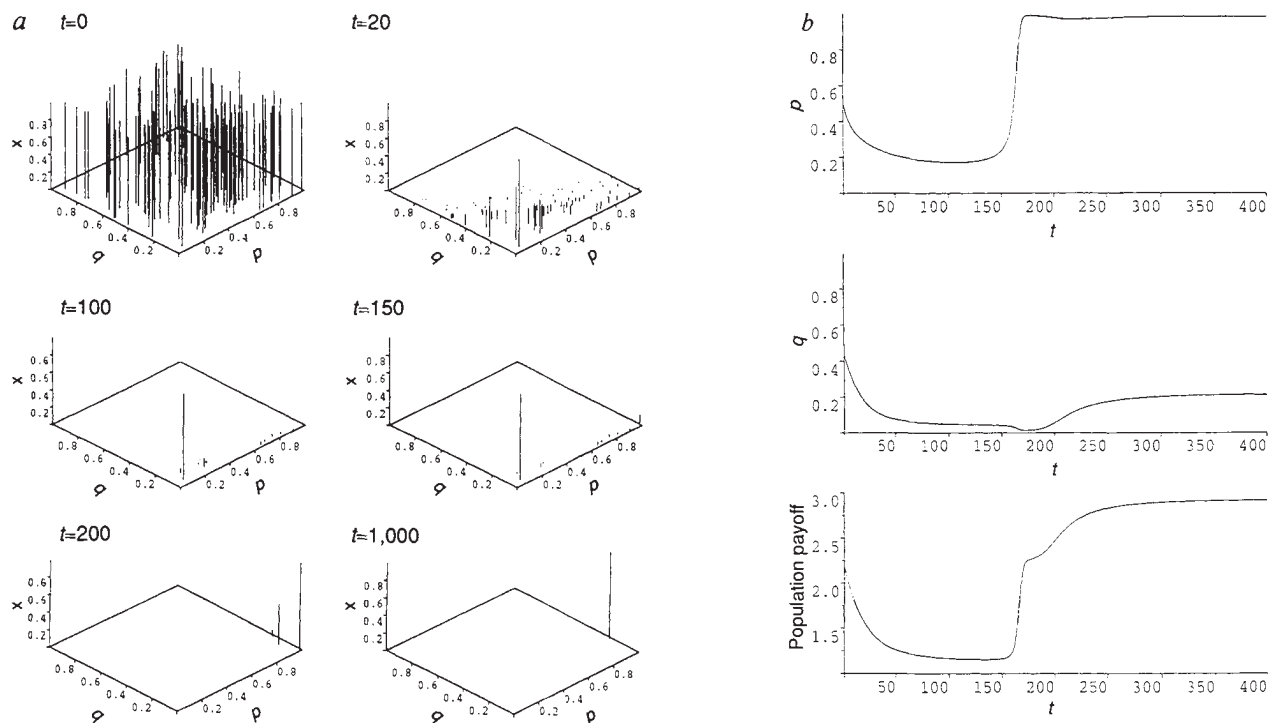


FIG. 1 The role of TFT for the emergence of cooperation in heterogeneous ensembles of strategies is illustrated in this simulation. The population consists of a randomly chosen sample of 99 reactive strategies $E_i = (p_i, q_i)$ uniformly distributed on the unit square. The stochastic TFT-like strategy $E = (0.99, 0.01)$ is then added to the population. The payoff for a strategy E_i against E_j in the infinitely IPD is obtained as $A(E_i, E_j) = 1 + 4c' - c - cc'$, with $c = [q_i + (p_i - q_i)q_j]/[1 - (p_i - q_i)(p_j - q_j)]$ and c' the analogous expression with i and j interchanged. Initially all strategies are present in the same frequency, 1%. If x_i denotes the frequency of E_i in one generation, then its frequency in the next generation will be given by $x'_i = x_i f_i(x)/\bar{f}$ for $i = 1, \dots, 100$. Here $f_i(x)$ denotes the average payoff for E_i in a population given by $x = (x_1 \dots x_{100})$ and $\bar{f} = \sum x_i f_i(x)$ is the average payoff in the population. During the first 100 generations the strategies close to ALLD

dominate the population. After about 150 generations stochastic TFT reappears and takes over. Reciprocity is established. This enables more forgiving strategies to grow. The final winner is GTFT ($p \approx 0.99$, $q \approx 0.33$). **a**, Relative frequencies (normalized such that the relative frequency of the most abundant strategy equals one) for all strategies after 0, 20, 100, 150, 200 and 1,000 generations. In our plot, strategies with very low frequencies are no longer visible, although they are still present in the numerical computations. **b**, Population averages for p , q and payoff. The apparition of stochastic TFT increases average payoff from about 1.2 to 2.25 (the payoff for the completely random strategy). To catalyse the turn of the tide, a strategy has to be very close to TFT. A strategy like (0.9, 0.1) will not, in general, be close enough to serve this purpose. Fully cooperative play (payoff 3) is only established after GTFT has taken over. Generosity pays under conditions of uncertainty.

best, whatever the other player does. Hence both players will use D and end up with one point only, instead of three points for mutual cooperation. This has nothing to do with rationality or foresight. We have in mind biological applications, where the payoff is number of offspring (more successful strategies have higher fitness) and strategies are inherited. The defectors, then, will necessarily out-compete the cooperators.

For the 'iterated prisoner's dilemma' (IPD), one assumes a constant probability w for another round; w can also be viewed as discount factor for future payoff. A strategy is a stochastic or deterministic rule specifying the choice of C or D in every round as a function of the history of the interaction so far.

In Axelrod's round-robin tournaments, TFT did best. It consists in playing C in the first round and from then on doing whatever the other player did in the previous round. Axelrod stressed, however, that TFT is not the best strategy (for $w > (T - R)/(T - P) = 2/4$, there is no best strategy in the IPD), and that in a pairwise interaction with any given partner, TFT never does better than this partner. But TFT can elicit cooperation from players who might defect among themselves.

There are essentially two ways of studying TFT's evolutionary role. One can check whether a homogeneous TFT population will resist invasion by mutant strategies. No single mutant can do better than TFT¹ if $w > \max((T - R)/(T - P), (T - R)/(R - S) = 2/3$. But ALLC (which always cooperates) and many other strategies do as well as TFT.

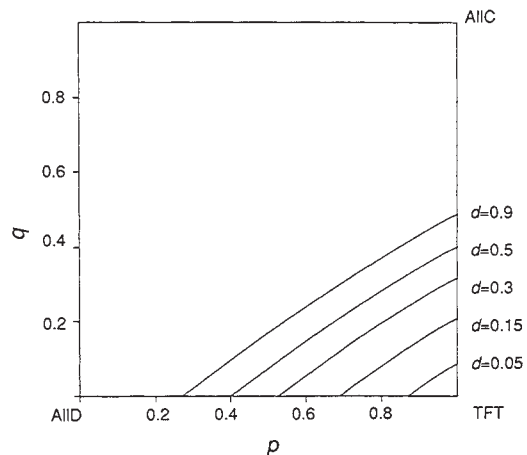
The heteromorphic approach starts with a widely scattered distribution and subjects it to selection. This has been done in

Axelrod's 'ecological' kind of tournament¹, when 63 entries were submitted, generation after generation, to a round-robin contest, their frequencies proportional to their payoff in the previous generation. The main problem here is to find plausible values from which to start. The participants in Axelrod's tournament included a good sample of game-theorists, but are unlikely to be representative of simpler biological communities.

The array of strategies for the IPD is so huge that it cannot be sampled by statistical means and has to be drastically reduced. Axelrod¹¹, for instance, considered deterministic strategies where the decision to cooperate or defect in each round depended on the outcome of the three previous rounds. Here, we consider a situation both simpler and more complex: the decision depends only on the previous round, but is stochastic and not deterministic. It reflects the tendency to cooperate in answer to the other's move. This is meant to apply to biological interactions where memories are short and decisions uncertain. Players may misinterpret the other's move (or identity), or misimplement their own intention. This relates not only to the 'trembling hand' behind the concept of perfect equilibrium^{12,13}, but also to the 'blurred minds'.

Errors are costly to a TFT population¹. Mistakes gives rise to sequences of alternating defections. This weakness of TFT is most pronounced in interactions against itself. When Axelrod repeated his tournament with an error rate of 1%, TFT still finished first, because the field of contestants was so diversified, but in a monomorphic TFT-population, the smallest error rate reduces fitness by 25%. It may be better to forgive with

FIG. 2 Strategies able to invade AllD in a cluster. For given cluster size d , when can an AllD population be invaded by a cluster of (p, q) strategists? If d is the frequency of (p, q) and $1-d$ that of AllD, the frequency of (p, q) grows if and only if $d > (1-p+q)^2/[3p-4q-3(p-q)^2]$. For a given d , this defines a neighbourhood of TFT in the unit square of strategies, drawn here for $d=0.9, 0.5, 0.3, 0.15$ and 0.05 . If d decreases, the neighbourhood shrinks to the corner point $(1, 0)$. Thus TFT is the strategy that can invade defectors in a minimal cluster.



a certain probability.

A reactive strategy is given by a triple (y, p, q) , where y is the probability to cooperate in the first round and p and q are the conditional probabilities to cooperate after a C (respectively D) of the other player¹⁴⁻¹⁶. TFT corresponds to $(1, 1, 0)$, AllD to $(0, 0, 0)$, AllC to $(1, 1, 1)$, 'suspicious tit-for-tat' (STFT) to $(0, 1, 0)$, and so on. We shall mostly be interested in properly stochastic strategies, with $0 < y, p, q < 1$.

For simplicity, we consider only the infinitely IPD (the limit case, $w = 1$). The initial move has no role, then, as its effect is 'forgotten' in the long run. A strategy now will simply be a point (p, q) in the unit square. The assumption $w = 1$ is only to keep things simple. The results still hold if we allow for some discount of the future ($w < 1$).

A simple characterization¹⁶ of when a strategy can be invaded by another strategy shows that 'generous tit-for-tat' (GTFT), the strategy $(1, q)$ with $q = \min\{1 - (T - R)/(R - S), (R - P)/(T - P)\} = \frac{1}{3}$, is optimal in the sense that among all reactive strategies immune to invasion by less cooperative strategies (lower p or q values), it affords the highest payoff for a population adopting it^{15,17}.

If we choose a representative sample of reactive strategies E_1 to E_n , initially all equally frequent, we can watch their frequencies evolve under the action of selection. With $n = 100$ different reactive strategies uniformly distributed on the unit square, evolution proceeds in most cases towards AllD: those (p, q) -strategies from the sample which are closest to $(0, 0)$ increase in frequency, while all others vanish. This follows because a large percentage of the random sample has high q values and does not retaliate against exploiters. With a rich diet of 'suckers', it pays to defect.

The outcome alters dramatically if one of the initial strategies (added by hand or by chance), is TFT, or very close to it (Fig. 1). The first phase is practically indistinguishable from the previous run. The strategies near AllD grow rapidly. TFT and all other reciprocating strategies (near $(1, 0)$) seem to have disappeared. But an embattled minority remains and fights back. The tide turns when 'suckers' are so decimated that exploiters can no longer feed on them. Slowly at first, but gathering momentum, the reciprocators come back, and the exploiters' now wane. But the TFT-like strategy that caused this reversal of fortune is not going to profit from it: having eliminated the exploiters, it is robbed of its mission and superseded by the strategy closest to GTFT. Evolution then stops. Even if we introduce occasionally 1% of another strategy, it will vanish.

Are reactive strategies a representative ensemble? They cover a broad range of cooperative, defective, fully random or reciprocal behaviour. They include strategies like AllC, AllD, TFT and GTFT. Their stochasticity reflects life's fuzziness. Their simplicity seems to match the metaphorical nature of the 'prisoner's dilemma'. That simple strategies are important for the IPD is shown by TFT's triumph in Axelrod's tournaments.

Will these findings hold for larger sets of strategies? Our simulations can be extended to more complex strategies and $w < 1$. Usually, however, the strategy space becomes so large that representative samples are too unwieldy. We therefore sketch only one extension, which covers stochastic strategies depending on the other player's last move, but also one's own. They are given by the four conditional probabilities p_1, p_2, p_3 and p_4 to cooperate after $(C, C), (C, D), (D, C)$ [respectively (D, D)] in the previous round. In these simulations behaviour is similar: strategies near TFT ($p_1 = p_3 = 1; p_2 = p_4 = 0$) favour the emergence of strategies near GTFT ($p_1 = p_3 = 1; p_2 = p_4 = 1/3$).

Axelrod¹¹ simulated the evolution of deterministic strategies depending on the previous three moves by both players (a promising variant including noise is described in ref. 18). Together with the specifications for the first rounds, such strategies are represented by strings of 70 C s and D s. Axelrod used a genetic algorithm, starting with a random sample of strategies, applying game dynamics and occasionally modifying strategies by mutating letters or recombining strings. Evolution always started off towards defection, but then veered towards cooperation. One can ask whether such a change is always due to the occurrence of a TFT-like strategy. Stochastic strategies like GTFT cannot occur here: but it would be interesting to know whether, here again, a generous strategy (for example, tit for two tats) would be the ultimate beneficiary of the action of stern retaliators. Or does forgiveness pay only under uncertainty? Axelrod and Dion⁴ suggest that for larger noise "generosity invites exploitation". In fact, the optimal forgiveness in reactive strategies decreases if noise grows¹⁵.

In our simulations, TFT is almost specified by its police role: strategies that are not very close to TFT do not have its effect. But an evolution twisted away from defection (and hence due to TFT) leads not to the prevalence of TFT, but towards more generosity. TFT's strictness is salutary for the community, but harms its own.

TFT acts as a catalyser. It is essential for starting the reaction towards cooperation. It needs to be present, initially, only in a tiny amount; in the intermediate phase, its concentration is high; but in the end, only a trace remains. □

Received 19 August; accepted 5 November 1991.

1. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
2. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390-1396 (1981).
3. Trivers, R. *Social Evolution* (Cummings, Menlo Park, 1985).
4. Axelrod, R. & Dion, D. *Science* **242**, 1385-1390 (1988).
5. Boyd, R. & Lorberbaum, J. P. *Nature* **327**, 58-59 (1987).
6. May, R. M. *Nature* **327**, 15-17 (1987).
7. Milinski, M. *Nature* **325**, 434-435 (1987).
8. Wilkinson, G. S. *Nature* **308**, 181-184 (1984).
9. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, UK, 1982).
10. Hofbauer, J. & Sigmund, K. *The Theory of Evolution and Dynamical Systems* (Cambridge University Press, UK, 1988).
11. Axelrod, R. in *Genetic Algorithms and Simulated Annealing* (ed. Davis, D.) (Pitman, London, 1987).
12. Selten, R. *Internat. J. Game Th.* **4**, 25-55 (1975).

13. Boyd, R. *J. theor. Biol.* **136**, 47–56 (1989).
14. Nowak, M. & Sigmund, K. *J. theor. Biol.* **137**, 21–26 (1989).
15. Nowak, M. & Sigmund, K. *Acta appl. Math.* **20**, 247–265 (1990).
16. Nowak, M. *Theor. Pop. Biol.* **38**, 93–112 (1990).
17. Molander, P. *J. Conflict Resolut.* **29**, 611–618 (1985).
18. Lindgren, K. in *Artificial Life II* (eds Farmer, D. et al.) (Proc. Santa Fe Inst. Stud., Addison-Wesley, 1991).

ACKNOWLEDGEMENTS. Support from the Austrian Forschungsförderungsfond is gratefully acknowledged.

Regulation of the G1–S transition in postembryonic neuronal precursors by axon ingrowth

Scott B. Selleck, Cayetano Gonzalez*, David M. Glover* & Kalpana White

Department of Biology, Brandeis University, Waltham, Massachusetts 02254-9110, USA

*Cancer Research Campaign Laboratories, Cell Cycle Genetics Group, Department of Biochemistry, The University, Dundee DD1 4HN, UK

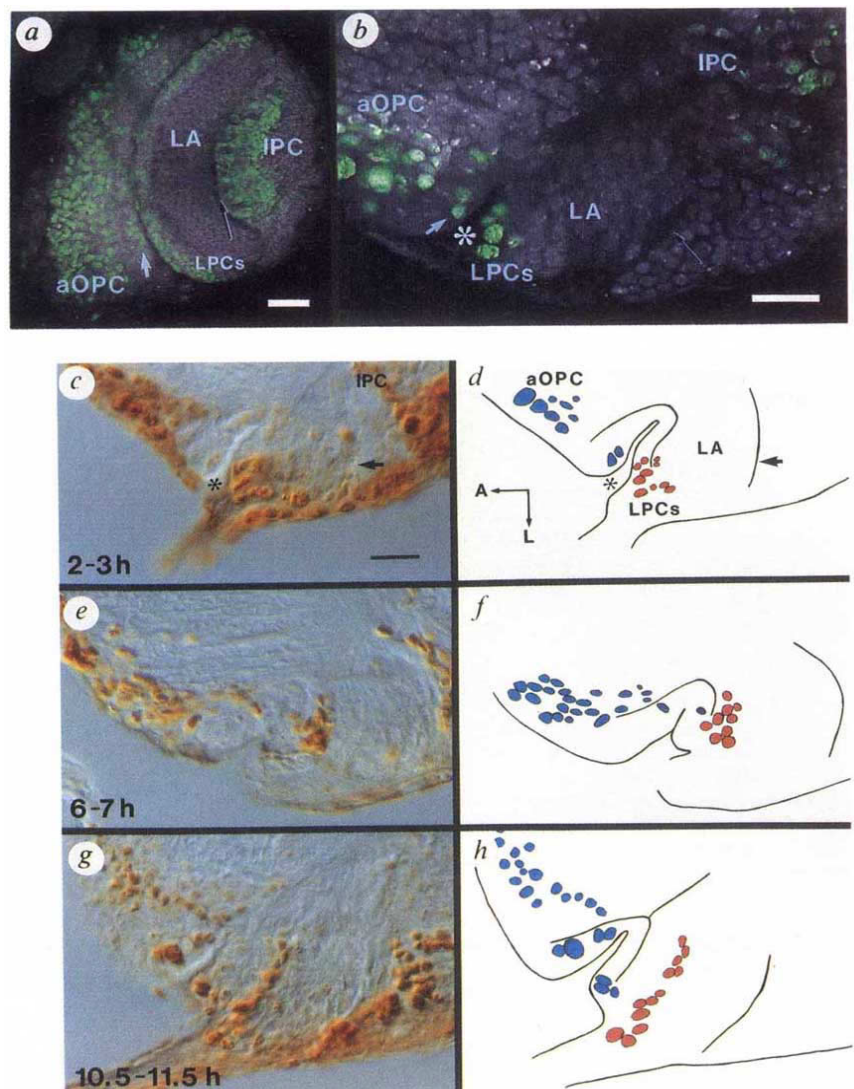
In the newly cellularized *Drosophila* embryo, progress through the cell cycle is regulated at the G2–M transition^{1,2}. We have examined cell-cycle regulation later in *Drosophila* development, in a group of postembryonic neuronal precursors. The S-phase precursor cells, which generate photoreceptor target neurons

(lamina neurons) in the central nervous system, are not present in the absence of photoreceptor innervation³. Here we report that axons selectively approach G1-phase precursors. Without axon ingrowth, lamina precursors do not enter their final S phase and by several criteria, arrest in the preceding G1 phase. These findings provide evidence that at this stage in development the control of cell division can occur at the G1–S transition.

Neurons of the adult optic lobe arise from cell divisions during postembryonic life⁴. Lamina neurons, which receive direct synaptic input from photoreceptors, are derived from a discrete group of S-phase lamina precursor cells (LPCs)^{3,5}. S-phase LPCs are part of a proliferative epithelium on the surface of the brain, the outer proliferative centre (OPC). The OPC epithelium invaginates anterior to the developing lamina to form a 'V'-shaped furrow (Fig. 1b, d). Pulse-labelling with bromodeoxyuridine (BUdR) shows three OPC S-phase domains. They are (from anterior to posterior): (1) the anterior OPC (aOPC), a broad belt of cells; (2) scattered BUdR-incorporating cells slightly more posterior; and (3) S-phase LPCs (Fig. 1a, b). To determine the origins and fate of S-phase LPCs, larvae were pulsed with BUdR and the labelled cells followed after varying lengths of time. Six hours after BUdR injection, cells labelled as S-phase LPCs (red cells in Fig. 1d, f, h) have moved into the lamina *en bloc* (compare Fig. 1c, d with e, f). By 12 h after BUdR delivery, labelled cells originating from the anterior region of the OPC (blue cells, Fig. 1d, f, h) have travelled along the furrow to reside where S-phase LPCs were seen with a pulse label (compare Fig. 1c, d with g, h). (aOPC progeny also become

FIG. 1 Anatomy of the outer proliferative centre, and origins of S-phase LPCs. a, b, S-phase cells in the third instar larval CNS visualized by Bromouridine deoxyribose (BUdR) incorporation (shown in green). Propidium iodide-stained nuclei shown in grey tones. Confocal images showing lateral (a, anterior left, ventral down) and horizontal (b, anterior left, lateral down) views. The OPC S-phase domains are: (1) the anterior OPC (aOPC); (2) scattered cells along the anterior segment of the furrow (white arrows a, b); and (3) S-phase LPCs. Asterisk indicates furrow; black arrows, posterior limit of the lamina (LA); IPC, inner proliferative centre. Scale bars, a: 25 μ m, b: 10 μ m. c–h, BUdR pulse-chase analysis of LPCs. Cells pulse-labelled with BUdR (c, d, 2–3 h post-BUdR injection) were followed at 6–7 h (e, f) and 10.5–11.5 h (g, h) post-labelling. c, e, g, Paraffin sections cut in horizontal plane (anterior left, lateral down). BUdR-labelled cells show brown staining. d, f, h, Representations of sections shown in c, e, g, respectively. Only cells that incorporate BUdR while part of the OPC, and OPC progeny are indicated. Cells labelled while in the S-phase domains anterior to S-phase LPCs are represented in blue, cells labelled as S-phase LPCs in red. BUdR incorporating cells of the IPC, the occasional cell in the lamina and perineural sheath are not indicated in the diagrams. Abbreviations and markings as for a, b. Scale bar, 10 μ m.

METHODS. Wild-type (Canton-S) CNS preparations shown in a, b were labelled for 0.5 h in Schneider's medium containing BUdR (ref. 14), and viewed with a Biorad MRC 600 confocal microscope. For the pulse-chase analysis, wild-type climbing third instar larvae were injected with BUdR (refs 3, 14) and incubated at 25 °C before dissection. BUdR-incorporating cells were detected using anti-BUdR antibody (Becton-Dickinson)^{3,14} and 6 μ m paraffin sections made¹⁵.



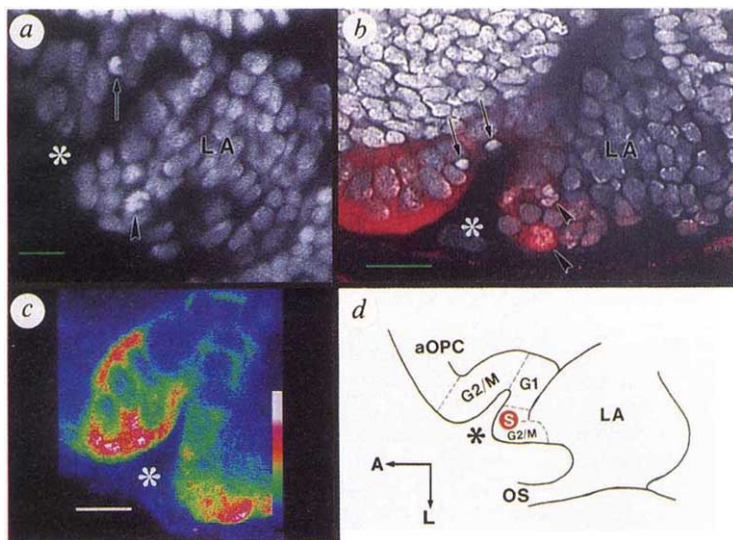


FIG. 2 Cell-cycle phase of LPCs along the OPC epithelium. *a*, *b*, confocal optical sections of propidium iodide and α -cyclin B (*b* only) antibody stained late-third instar larval CNS preparations (anterior forward, lateral down). Mitotic chromosomes are seen as condensed structures with propidium staining (grey tones). M-phase cells are found along the anterior segment of the furrow (long arrows), and in the anterior portion of the lamina (arrowheads). *b*, Cyclin B-IR represented in red. Cyclin A-IR (not shown) follows a similar pattern. Asterisk marks the furrow. Scale bars, 10 μ m. *c*, Cyclin B-IR along OPC furrow; signal intensities are represented on colour scale (white high, blue low; see intensity wedge). *d*, Diagram of OPC epithelium and lamina showing cell-cycle progression of LPCs. M-phase domains were assigned on the basis of greater than 200 mitotic figures counted from more than 20 serially sectioned preparations. LA, Lamina; OS, optic stalk. METHODS. CNS-eye disc complexes were dissected from wild-type (Canton-S) climbing third instar larvae. Propidium iodide and anti-cyclin B antibody staining as in ref. 7.

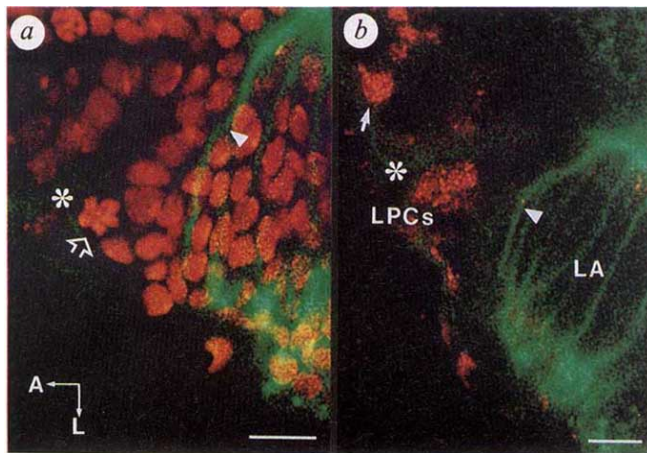
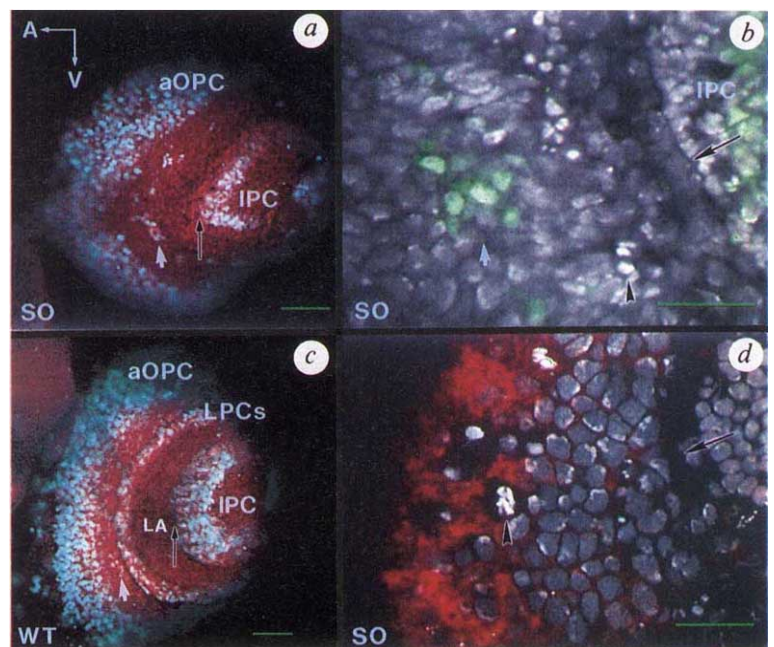


FIG. 3 Location of photoreceptor axons relative to LPCs. *a*, Position of photoreceptor axons (in green, stained with α -HRP antibody), and propidium-stained nuclei (red) seen in horizontal optical section (anterior left, lateral down). The most anterior axon bundles (arrowhead) run adjacent to G1-phase LPCs, along the furrow (furrow marked with asterisk). Axons are some distance from the lamina mitotic domain (open arrow points to a mitotic cell). *b*, Location of photoreceptor axons (green) relative to S-phase LPCs (red). Note that S-phase LPCs are not in contact with the most anterior axon bundle (arrowhead). An S-phase cell along the anterior segment of the furrow is marked with a white arrow (see Fig. 1*a, b*). LA, lamina. Scale bars, 10 μ m.

METHODS. Photoreceptor axons visualized using α -HRP antibody conjugated to FITC (Cappel), according to published protocols^{3,16}. BUdR incorporation and propidium iodide staining was done as described above.

FIG. 4 Cell-cycle progression of LPCs in the absence of innervation. *a*, *c*, S-phase cells in wild-type (*c*) and *sine oculis* (*so*) (*a*) larval brains. Lateral views (anterior left, ventral down) showing BUdR-incorporating cells (blue) and propidium-stained nuclei (red). Note the absence of S-phase LPCs in *so* brain. The two domains of S-phase cells normally anterior to S-phase LPCs are seen in *so* (aOPC, and scattered cells more posterior; white arrow). Long arrows mark anterior margin of IPC. There is no discernible lamina (LA) in *so*. Scale bar, 25 μ m. *b*, *so* CNS (lateral view) stained with anti-BUdR antibody (green) and propidium iodide (grey tones). As in the wild type, M-phase cells (one indicated by black arrowhead) are found posterior (to the right) to the S-phase domain seen as scattered BUdR incorporating cells (white arrow). No BUdR-incorporating cells are seen posterior to M-phase domain. In the absence of photoreceptor axon ingrowth, OPC progeny ultimately reach the boundary with the IPC (long arrow) (data not shown) but do not differentiate into lamina neurons³. Scale bar, 10 μ m. *d*, *so* CNS stained with α -cyclin B antibody (red) and propidium iodide (grey tones), lateral view. High cyclin B-IR marks late G2 domain located anterior to M-phase domain (arrowhead points to one mitotic figure). Interphase cells posterior to M-domain show low levels of cyclin B-IR. Not all innervation minus *so* brains show as low a level of cyclin B-IR in posterior region of epithelium as shown in this preparation, although the zones of high followed by low cyclin B-IR are consistently observed. Scale bar, 10 μ m.

METHODS. *so* larvae show variable expressivity for the number of photoreceptor cells which develop¹⁰. Staining of *so* third instar larvae with α -HRP antibody to visualize photoreceptor projections, and propidium iodide, has shown that in the absence of innervation no lamina components form (data not shown and ref. 3). Lamina structure



seen by propidium iodide staining was used to identify larvae without photoreceptor innervation.

medullary neurons, seen medial to the aOPC (Fig. 1e-h).^{4,5}. This analysis indicates that S-phase LPCs are continuously produced from dividing cells in the anterior region of the OPC and move through the cell cycle as they proceed toward the lamina. Cells along the length of the OPC epithelium therefore represent successive cell-cycle phases.

We have further examined the cell-cycle phases of LPCs as they proceed into the lamina. *Drosophila* cyclin A- and B-immunoreactivity (IR) peak in late G2 to early M, and show a rapid decline by anaphase⁶⁻⁸. Although the timing of cyclin reaccumulation is not precisely known, early G1 cells show low levels of cyclin-IR compared with late G2 stages⁶⁻⁸. LPCs show high cyclin B-IR levels (marking late G2-phase cells) in two regions (Fig. 2b, c). One of these regions precedes and overlaps with a domain of M-phase cells along the furrow (long arrows, Fig. 2a, b). These mitotic cells are anterior to the S-phase LPCs (see Figs. 1b and 2d). Thus cells located between this M-phase domain and the S-phase LPCs must be in G1. S-phase LPCs then proceed through the cycle into G2. This is seen as the second peak of cyclin B-IR (Fig. 2b, c), which precedes mitosis in the lamina itself (arrowheads, Fig. 2a, b). Thus cells enter the furrow in G2, divide and then complete one more cell cycle on their way into the lamina (Fig. 2d).

Using confocal microscopy, we have determined which cells of the proliferative epithelium contact photoreceptor axons (Fig. 3). The most anterior axon bundles (the latest to arrive from the developing eye) are found adjacent to G1-phase LPCs in the posterior segment of the furrow (Fig. 3a). Axons are some distance away from the S-phase LPCs (Fig. 3b), and M-phase cells in the lamina (Fig. 3a). The location of axon bundles suggests that the signal to proceed through the cell cycle is delivered to G1-phase LPCs.

If photoreceptors deliver a signal needed by LPCs to enter S phase, then in the absence of innervation, progenitors should not proceed beyond G1. This is observed in larvae lacking photoreceptor projections as a consequence of the mutation *sine oculis(so)*⁹. A BUdR pulse-chase analysis of larvae lacking innervation indicates that like the wild type, OPC progeny occupy progressively more posterior positions with time (data not shown). (The *so*⁺ gene product is not required in lamina neurons for their normal development^{10,11}.) The OPC epithelium of these brains lacks the final division cycle of LPCs seen in the wild type. G2 cells with high levels of cyclin B-IR are followed by an arc of mitotic cells (Fig. 4d). That LPCs then progress into G1 is evidenced by the presence of interphase cells in the most posterior, and hence most mature segment of the epithelium, which expresses low levels of cyclin B-IR (Fig. 4d). S-phase LPCs are not seen (compare *so* central nervous system (CNS) in Fig. 4a, b with wild type in c). This division cycle is apparently blocked in G1, and a lamina does not develop. We do not know the ultimate fate of cells produced from the OPC that would normally become lamina neurons. The lack of significant accumulation of cells suggests that in the absence of innervation G1-phase LPCs ultimately degenerate.

Several recent studies have documented the interaction between peripheral organs and neurogenesis in the CNS^{3,12}. Studies using an *in vitro* vertebrate system show that neuronal processes can stimulate cell division by direct contact¹³. We have provided evidence that in a group of postembryonic cells in the *Drosophila* CNS, cell-cycle progression is regulated at the G1-S boundary. This contrasts with the control of the cell cycle in G2 for early embryonic divisions^{1,2}. The presence of photoreceptor axons adjacent to G1-phase LPCs suggests that contact is required for the transmission of the mitogenic signal. □

4. White, K. & Kankel, D. R. *Dev. Biol.* **65**, 296-321 (1978).
5. Hofbauer, A. & Campos-Ortega, J. A. *Roux's Arch. dev. Biol.* **198**, 264-274 (1990).
6. Lehner, C. F. & O'Farrell, P. H. *Cell* **56**, 957-968 (1989).
7. Whitfield, W. G. F., Gonzalez, C., Maldonado-Codina, G. & Glover, D. M. *EMBO J.* **9**, 2563-2572 (1990).
8. Lehner, C. F. & O'Farrell, P. H. *Cell* **61**, 535-547 (1990).
9. Lindsley, D. L. & Grell, E. H. *Genetic Variations of Drosophila melanogaster* (Carnegie Inst. Publ. No. 627, Washington, D.C., 1968).
10. Fischbach, K. F. *Dev. Biol.* **95**, 1-18 (1983).
11. Fischbach, K. F. & Technau, G. *Dev. Biol.* **104**, 219-239 (1984).
12. Baptista, C. A., Gershon, T. R. & Macagno, E. R. *Nature* **346**, 855-858 (1990).
13. Ratner, N., Hong, D., Lieberman, M., Bunge, R. & Glaser, L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6992-6996 (1988).
14. Truman, J. W. & Bate, M. *Dev. Biol.* **125**, 145-157 (1988).
15. White, K., Hurteau, T. & Punsal, P. J. *comp. Neurol.* **247**, 430-438 (1986).
16. Jan, L. Y. & Jan, Y. N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2700-2704 (1982).

ACKNOWLEDGEMENTS. We thank L. Restifo, T. Tully, B. Thomas and members of the White Laboratory for reading the manuscript. This work was supported by the NIH (to K.W.), CRC (to D.M.G. and C.G.) and a Burroughs Wellcome travel grant for US-Great Britain Exchange (to S.B.S.). S.B.S. is a Burroughs Wellcome Fund fellow of the Life Sciences Research Foundation.

Activation of a *C. elegans* *Antennapedia* homologue in migrating cells controls their direction of migration

Stephen J. Salser & Cynthia Kenyon

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0554, USA

ANTERIOR-POSTERIOR patterning in insects, vertebrates and nematodes involves members of conserved *Antennapedia*-class homeobox gene clusters (HOM-C) that are thought to give specific body regions their identities¹⁻⁵. The effects of these genes on region-specific body structures have been described extensively, particularly in *Drosophila*, but little is known about how HOM-C genes affect the behaviours of cells that migrate into their domains of function. In *Caenorhabditis elegans*, the *Antennapedia*-like HOM-C gene *mab-5* not only specifies postembryonic fates of cells in a posterior body region, but also influences the migration of mesodermal and neural cells that move through this region⁵⁻⁷. Here we show that as one neuroblast migrates into this posterior region, it switches on *mab-5* gene expression; *mab-5* then acts as a developmental switch to control the migratory behaviour of the neuroblast descendants. HOM-C genes can therefore not only direct region-specific patterns of cell division and differentiation, but can also act within migrating cells to programme region-specific migratory behaviour.

The migratory neuroblasts QL and QR are bilateral homologues born in left-right symmetrical positions on the sides of the animal. These cells divide in identical patterns to produce a mechanosensory neuron, an interneuron, a ciliated neuron, and two programmed cell deaths (Fig. 1a)^{8,9}. Curiously, the Q cells migrate in opposite directions¹⁰. Soon after the animal hatches, QR migrates anteriorly whereas QL migrates posteriorly (Fig. 1b). The descendants of the Q cells then continue migrating asymmetrically, to anterior positions on the right side and posterior positions on the left side (Fig. 1b).

Previously it has been shown that in *mab-5(lf)* mutants, the initial direction of Q migration is still bilaterally asymmetric, but at some point in the lineage there is a reversal in the direction of cell migration, causing the descendants of QL to adopt anterior positions¹¹. Although *mab-5* affects migration in this lineage, it does not seem to affect other aspects of differentiation, such as the ability of specific descendants to differentiate as mechanosensory neurons¹¹.

To determine the precise role of *mab-5* in these migrations, we observed the movements of each cell in the Q lineage in animals carrying putative null and gain-of-function alleles of *mab-5*, and also examined the effect of expressing *mab-5* under the control of a *C. elegans* heat-shock promoter.

Received 23 May; accepted 15 October 1991.

1. Edgar, B. A. & O'Farrell, P. H. *Cell* **57**, 177-187 (1989).
2. Edgar, B. A. & O'Farrell, P. H. *Cell* **62**, 469-480 (1990).
3. Selleck, S. B. & Steller, H. *Neuron* **6**, 83-99 (1991).

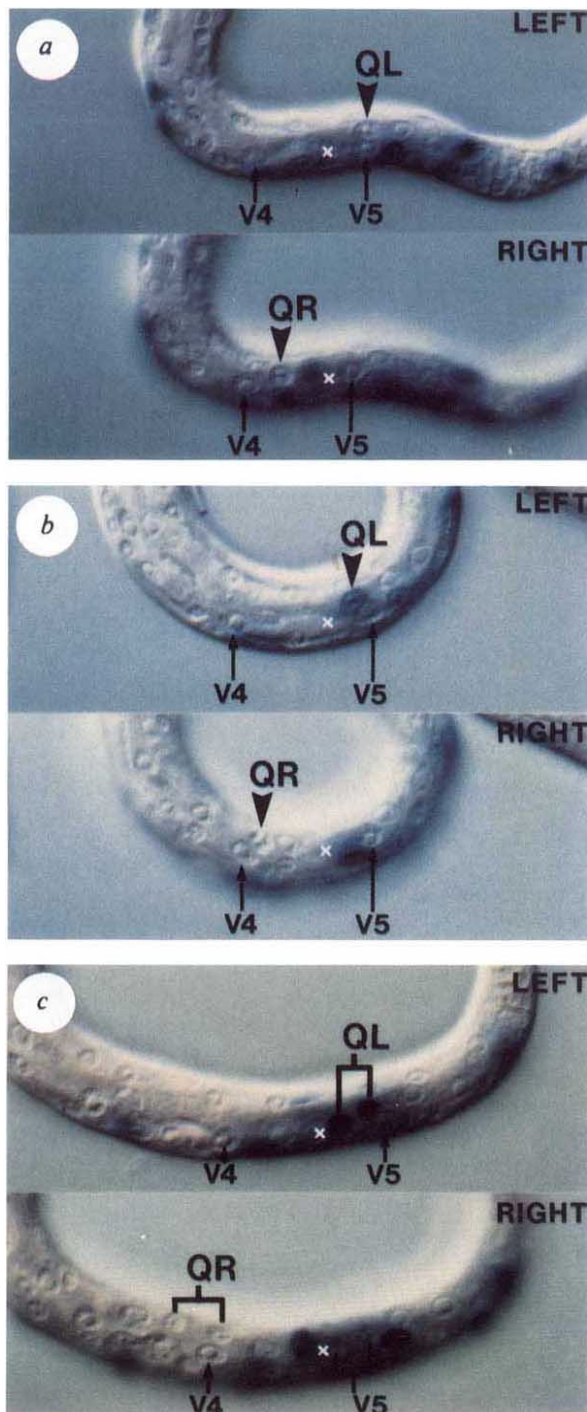


FIG. 2 Expression of the *mab-5-lacZ* fusion in Q cells and their descendants. The position of the Q cells at hatching is indicated by a white cross. *a*, Q cells begin migrating; neither cell expresses the fusion. *b*, Initial migrations of the Q cells are almost complete. QL expresses low levels of the fusion; none is detectable in QR. *c*, Q cells have just divided. QL descendants express high levels of the fusion; none is detectable in QR descendants. Staining persists in all QL descendants throughout their migration and to the end of L1. This pattern is highly reproducible: 0–2 h after hatching, no expression is seen in either Q cell ($n=38/38$ animals). At 2–3 h, some expression is observed in QL cells migrating up over V5 ($n=8/14$); no expression is seen in QR ($n=14/14$). After Q division, strong expression is observed in QL but not QR descendants ($n=20/20$).

METHODS. Synchronous *unc-31(e169)*; *muls2; e1490* populations were obtained by washing newly hatched worms off plates; *muls2* is a chromosomal integration containing the *mab-5-lacZ* construct and the cotransformation marker *unc-31*. (An independent integrant constructed by D. Cowing and cotransformed with *rol-6* behaves similarly.) The construct contains the first 17 amino acids of *mab-5* and 7 kb of upstream sequences fused in-frame to the *lacZ* expression vector pPD21.28 (ref. 27), containing the SV40 nuclear localization signal. These upstream sequences when joined to the rest of *mab-5* are sufficient to produce phenotypic rescue in all tissues known to be affected by *mab-5*. The fusion is expressed in all the posterior cells in which *mab-5* function is predicted genetically, as well as the posterior intestine and juvenile neurons. Larvae were stained using a variation of the method described in ref. 27.

In all of these strains, the asymmetric migrations of the Q cells themselves are normal (Fig. 1, *b–e*). As these migrations are unaffected by *mab-5(e2088lf)*, as well as several other putative null mutations (*e1239* (refs 7, 11); *e1936*, *e2011*, *n1384* (data not shown)), we believe that the initial asymmetry in the Q migrations is independent of *mab-5*. Furthermore, neither the gain-of-function allele, nor induction of heat-shock-*mab-5* expression before Q migration ($n=19$) perturbs this migration, suggesting that the migration of Q itself is not only independent of, but also immune to, *mab-5* activity.

Changes in *mab-5* activity do, however, strongly affect the migratory behaviours of the Q daughters, Q.a and Q.p, and their descendants. In a *mab-5(e2088lf)* background, these cells migrate anteriorly on both sides (Fig. 1c), as seen on the wild-type right side. This shows that *mab-5(+)* activity is necessary to prevent the QL descendants from migrating anteriorly.

To test whether *mab-5* would also be sufficient to make the QR descendants stay in the posterior, we expressed *mab-5* under the control of a *C. elegans* heat-shock promoter in a *mab-5(lf)* background (Fig. 1d). This produced a phenotype reciprocal to *mab-5(lf)* mutations: Q.a and Q.ap migrated posteriorly on both sides, whereas Q.p and its descendants simply stayed put as they normally do on the left side of the wild type.

As an independent confirmation that *mab-5* expression is sufficient to make the QR descendants stay in the posterior, we examined animals carrying a putative *mab-5* gain-of-function

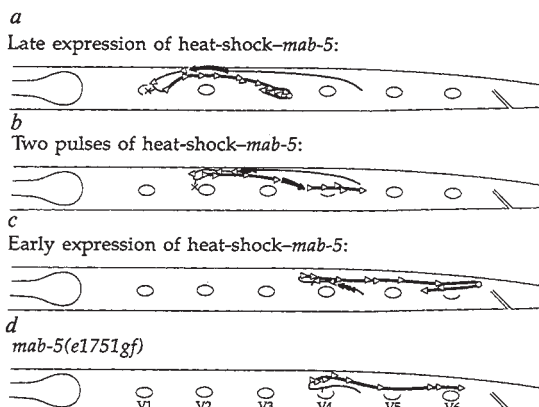


FIG. 3 Reversal of migration using a pulse of *mab-5*. *a*, Before heat-shock-*mab-5* expression was introduced, Q.a migrated anteriorly (initial migration of Q and Q.a is shown by the black arrow; the period of heat-shock is indicated in red). After heat-shock, Q.a continued migrating anteriorly, divided, and its descendant Q.ap migrated posteriorly (in bold). Open arrowheads indicate datapoints and inferred direction of migration. In two other animals tested, behaviour was similar. In all cases Q.ap turned around a second time, about 4 h after the heat-shock, and migrated a short distance anteriorly before stopping. In non-heat-shocked controls ($n=10$) or animals heat-shocked but lacking the construct ($n=11$), Q.a and Q.ap continued migrating to the head. *b*, Q.ap was prevented from turning around a second time with a second pulse of *mab-5* (second red arrow). Similar results were seen in 6/6 animals. *c*, As described in the legend to Fig. 1, when heat shock is used to turn on *mab-5* before the division of Q, both Q.a and Q.ap migrate posteriorly. In those animals in which Q.ap failed to reach the tail by ~4 h after heat shock, it turned around again and migrated anteriorly ($n=2$). *d*, Migration path of Q.a and Q.ap in *mab-5(e1751gf)*. Q.a migrates posteriorly to a position behind V6 and does not turn around even though it fails to reach the tail ($n=5/5$). For all the animals described in this figure, the time at which Q.ap stops migrating is reproducible (138 ± 23 min after Q.a division, range 94–176 min.).

allele, *e1751* (ref. 12). The effects of *e1751* are similar to those seen with the heat-shock construct (Fig. 1e). In addition, *e1751* maps close to *mab-5* (E. Hedgecock, personal communication). We have analysed the mutation and found that it consists of a tandem duplication of the entire *mab-5* gene, with one duplication breakpoint 4 kilobases (kb) upstream of *mab-5*. The gain-of-function phenotype does not result simply from the doubled dosage of *mab-5* because it cannot be corrected by placing *e1751* over a genetic deficiency spanning the region (thus restoring the normal number of gene copies). Instead it seems likely that the duplication breakpoint alters transcription of the neighbouring *mab-5* gene, leading to increased or ectopic expression. From these results, together with the heat-shock experiments and the *e2088lf* phenotype, we conclude that *mab-5* transcription is both necessary and sufficient to give Q descendants the identities they would have as QL descendants rather than QR descendants.

It is interesting that *mab-5* expression has different effects on different cells within the lineage: Q.a and its daughter respond to *mab-5* by migrating posteriorly rather than anteriorly. Q.p and its daughters respond to *mab-5* by staying put instead of migrating anteriorly. The difference in the way these cells respond to *mab-5* does not seem to depend critically on the dosage or timing of *mab-5* expression, as the same effects are seen in response to *e1751gf*, the heat-shock-*mab-5* construct, and wild-type activity on the left side. It seems likely that the different descendants are intrinsically programmed to respond in different ways to *mab-5*.

Where is *mab-5* expressed? Previous mosaic analysis suggested that *mab-5* acts autonomously in migrating mesodermal and neural cells⁷. If so, then as *mab-5* transcription seems to be both necessary and sufficient to generate the pattern of migration normally seen on only the left side, it follows that this transcription should be restricted to cells of the left, but not the right, Q lineage. To test this model, we constructed a *mab-5-lacZ* fusion and generated several transgenic lines (Fig. 2). In these lines, the construct is expressed in the posterior of the worm in a pattern consistent with earlier *in situ* analysis⁶.

We find that *mab-5-lacZ* is expressed in precisely those cells of the Q lineage for which its function was demonstrated genetically. First, expression is specific for the left Q-cell lineage; the fusion is not expressed in QR or its descendants. Furthermore, *mab-5-lacZ* is not expressed in QL during the initial *mab-5*-independent phase of its migration, but is switched on in QL during the course of its migration (Fig. 2b). The β -galactosidase staining becomes most intense after the first division and persists throughout the migration of the QL descendants (Fig. 2c). This pattern of expression confirms the predictions of the mosaic and genetic analysis, and argues strongly that *mab-5* indeed acts within the QL lineage to programme cell migrations.

What kind of guidance system directs these cells? For the posterior migrations of QL.a and QL.ap, *mab-5* presumably activates downstream cell migration genes that permit these cells to migrate posteriorly in response to some guidance cue. This cue could be a short-range attractant or it could be part of a global coordinate system that allows migrating cells to sense which direction is posterior, no matter where they are located. To distinguish between these possibilities, we allowed Q.a to migrate into the anterior in the absence of *mab-5*(+) activity, and then induced *mab-5* expression from the heat-shock promoter. This later expression of *mab-5* causes Q.ap to migrate posteriorly even when it is located far in the anterior (Fig. 3a and b). This indicates that the signals used to guide cells posteriorly are distributed along the entire body axis.

Interestingly, in these animals the migrating Q.ap cell turned around once again and began to migrate anteriorly about four hours after the heat shock (Fig. 3a). This behaviour was also seen in animals heat-shocked at an earlier stage (Fig. 3c). In contrast, the Q.ap cells did not turn around in animals receiving a second heat shock to maintain *mab-5* expression, or in animals carrying the gain-of-function allele of *mab-5* (Fig. 3b and d).

In these cases, the Q.ap cell simply continued migrating posteriorly until it stopped. These results suggest that the migratory behaviour of these cells is plastic and can be reversed by either providing or removing *mab-5*.

This study indicates that in addition to controlling region-specific cell division and terminal differentiation, HOM-C genes can also function in migrating cells to programme their behaviour in specific body regions. In fact, the only apparent feature of the Q descendants that is affected by *mab-5* is their migration^{8,11}. In vertebrates, HOM-C genes are expressed in migratory neural crest cells^{13,14} and can influence their development¹⁵⁻¹⁸. These genes may be involved in specifying the overall developmental fate of specific neural crest cells but, like *mab-5* in the Q lineage, they could also have a more specific role in directing cell migration.

The observed pattern of *mab-5-lacZ* expression implies that *mab-5* transcription is activated specifically in QL, as this cell

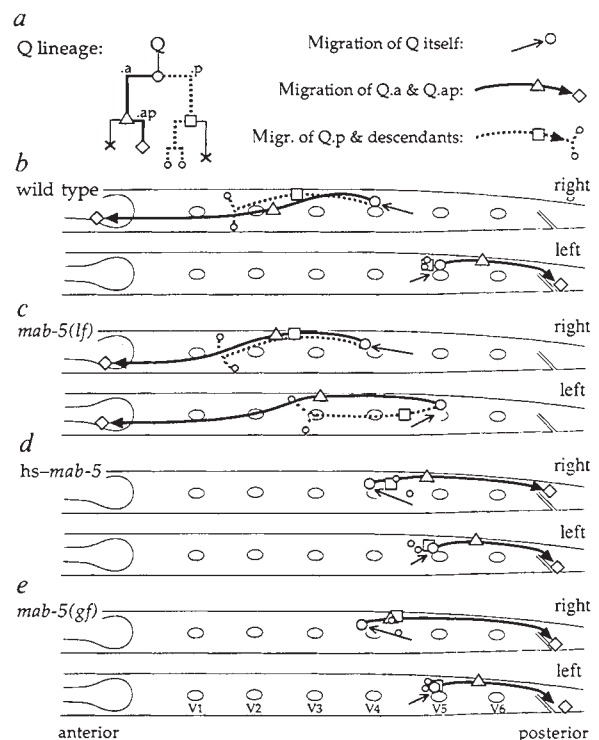


FIG. 1 Q cell migration in wild type and *mab-5* mutants: a, Q neuroblast lineage. b-e, Representative left and right side migrations in wild type, *mab-5*(*e2088lf*), *mab-5*(*e1751gf*), and animals expressing heat-shock-*mab-5*.

METHODS. Each drawing shows cell migration in representative animals observed using differential interference contrast optics. Complete cell migrations were observed in other animals and are reproducible. Genotypes of the strains examined were *e1490* ($n=5L, 5R$), *e2088*; *e1490* ($n=4L, 4R$), *e1751*; *e1490* ($n=4L, 5R$) and *e2088*; *unc-31*(*e169*); *e1490*; *muEx14* ($n=8L, 3R$). We used *him-5*(*e1490*) to increase the frequency of male progeny; no significant differences were seen between sexes. The extrachromosomal array *muEx14* carries the *hsp16.1-mab-5* construct. It was obtained according to ref. 23 using the cotransformation marker *unc-31* (cosmid C14G10; R. Hoskins, personal communication). An independent line cotransformed with *rol-6* (ref. 24) behaves similarly. The *hsp16.1/48* locus was provided by E. P. M. Candido^{25,26}. A *mab-5* complementary DNA was inserted in-frame into the second exon of *hsp16.1*, creating a fusion protein with the first 44 amino acids of *hsp16.1* and the last 183 amino acids of *mab-5*. Seventeen amino acids are removed from the N terminus of *mab-5*. To heat-shock, animals were incubated at 31 °C for 30 min after the initial Q migration but before division of Q. Migration in control animals receiving heat shocks without carrying the construct ($n=11$), or carrying the construct without receiving heat shocks ($n=10$), are indistinguishable from *mab-5*(*lf*). Heat shocks later in development produced a *mab-5*(*gf*) phenotype in other tissues in which *mab-5* is known to act.

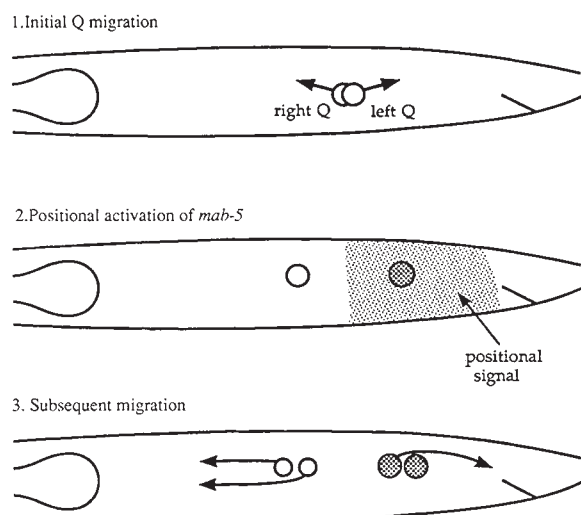


FIG. 4 Positional model for *mab-5* activation and function in the Q lineage.

migrates posteriorly into the general body region where *mab-5* is expressed and determines cell fate. QL moves into a region of the body where ectodermal, mesodermal and endodermal cells already express *mab-5-lacZ*, whereas QR migrates anteriorly into a region where only the juvenile neurons express *mab-5-lacZ*. One attractive explanation of the activation of expression in QL is that this cell encounters position-specific signals which activate *mab-5* transcription as it migrates into the posterior (Fig. 4). Alternatively, QR may avoid such signals by migrating anteriorly. Consistent with this model, when laser microbeam damage is used to block the initial anterior migration of QR, in some animals the QR descendants change their behaviour and migrate like wild-type QL descendants, as judged by their final positions in the posterior body region (J. Austin and C.K., unpublished results). In vertebrate development, it has been proposed that HOM-C gene expression can be activated by homeogenetic induction from adjacent tissues that already express that HOM-C gene¹⁶. However, the signals that activate *mab-5* expression in QL must be *mab-5*-independent, for expression of the *mab-5-lacZ* fusion is normal in a *mab-5*(-) background ($n = 6/6$ e2088; e1490 animals).

What about other HOM-C genes? In many systems, migrating cells and growth cones show complex pathfinding behaviours and change the expression of cell-surface molecules during migration¹⁹⁻²². In some cases it may be that as migrating cells cross from one spatial domain into another, they change their HOM gene expression to match their new surroundings. Subsequent changes in downstream gene expression could, in turn, reprogramme their migratory behaviour. In principle, such a system would be an effective means of guiding long-range cell migration. □

Received 29 July; accepted 29 October 1991.

- Gehring, W. J. *Science* **236**, 1245-1252 (1987).
- Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E. & Krumlauf, R. *Nature* **341**, 405-409 (1989).
- Reid, L. *Cell* **63**, 875-882 (1990).
- Bürglin, T. R. et al. *Nature* **351**, 703 (1991).
- Kenyon, C. & Wang, B. *Science* **253**, 516-517 (1991).
- Costa, M., Weir, M., Coulson, A., Sulston, J. & Kenyon, C. *Cell* **55**, 747-756 (1988).
- Kenyon, C. *Cell* **46**, 477-487 (1986).
- Chalfie, M. & Sulston, J. *Dev. Biol.* **82**, 358-370 (1981).
- White, J., Southgate, E., Thomson, J. & Brenner, S. *Phil. Trans. R. Soc.* **314B**, 1-340 (1986).
- Sulston, J. & Horvitz, H. *Dev. Biol.* **56**, 110-156 (1977).
- Chalfie, M., Thomson, J. & Sulston, J. *Science* **221**, 61-63 (1983).
- Hedgecock, E., Culotti, J., Hall, D. & Stern, B. *Development* **100**, 365-382 (1987).
- Frohman, M. A., Boyle, M. & Martin, G. R. *Development* **110**, 589-607 (1990).
- Hunt, P., Wilkinson, D. & Krumlauf, R. *Development* **112**, 43-50 (1991).
- Wolgemuth, D. J., Behringer, R. R., Mostoller, M. P., Brinster, R. L. & Palmiter, R. D. *Nature* **337**, 464-467 (1989).
- Cho, K. W. Y. et al. *EMBO J.* **7**, 2139-2149 (1988).
- Balling, R., Mutter, G., Gruss, P. & Kessel, M. *Cell* **58**, 337-347 (1989).

- Chisaka, O. & Capecchi, M. R. *Nature* **350**, 473-479 (1991).
- Damsky, C. H., Fitzgerald, M. L. & Fisher, S. J. *J. clin. Invest.* (in the press).
- Klamt, C., Jacobs, J. R. & Goodman, C. S. *Cell* **64**, 801-815 (1991).
- Elkins, T., Zinn, K., McAllister, L., Hoffmann, F. M. & Goodman, C. S. *Cell* **60**, 565-575 (1990).
- Dodd, J. & Jessell, T. M. *Science* **242**, 692-699 (1988).
- Fire, A. *EMBO J.* **5**, 2673-2680 (1986).
- Mello, C., Kramer, J., Stinchcomb, D. & Ambros, V. *EMBO J.* (in the press).
- Russnak, R. & Candido, E. P. M. *Molec. cell. Biol.* **5**, 1268-1278 (1985).
- Stringham, E. G., Dixon, D. K., Jones, D. & Candido, E. P. M. *Molec. Biol. Cell.* (in the press).
- Fire, A., White-Harrison, S. & Dixon, D. *Gene* **93**, 189-198 (1990).

ACKNOWLEDGEMENTS. We thank M. Tessier-Lavigne, G. Martin, M. Frohman, B. Wold, C. Bargmann and members of the Kenyon laboratory for discussion and comments on the manuscript; E. Hedgecock, R. Hoskins, A. Chisholm, D. Thierry-Mieg and N. T. Robinson for unpublished information on *mab-5*(e1751); E. Stringham and E. P. M. Candido for the *C. elegans* heat-shock promoter and for discussion; A. Fire for advice and for plasmid vectors; and E. Hedgecock, M. Stern, R. Horvitz, K.-L. Chow and S. Emmons for *mab-5* mutant strains. Some nematode strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. S.S. was supported by an NSF predoctoral fellowship and by a UCSF Chancellor's Award. C.K. is a Packard Foundation Fellow. This work was supported by a grant from the NIH.

Immunization of hu-PBL-SCID mice and the rescue of human monoclonal Fab fragments through combinatorial libraries

Michel A. Duchosal*, Sabine A. Eming*, Peter Fischer*†, Didier Leturcq‡, Carlos F. Barbas III†, Patricia J. McConahey*, Roger H. Caothien†, George B. Thornton‡, Frank J. Dixon* & Dennis R. Burton*†§

* Department of Immunology and † Molecular Biology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA
‡ RW Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, California 92121, USA
§ Krebs Institute, Department of Molecular Biology and Biotechnology, The University, Sheffield S10 2TN, UK

ANTIBODIES are usually prepared from recently boosted animals and reflect ongoing immune responses¹⁻⁶. In humans, this is restrictive as ethical constraints generally prevent antigen-boosting. Therefore the rich memory compartment of human antibody responses remains largely untapped⁷. Severe combined immune deficiency (SCID) mice⁸ populated with human cells⁹⁻¹¹ allow the stimulation of human antibody memory without the usual constraints. Here we show how peripheral blood lymphocytes can be stimulated by antigen to produce large secondary responses after transfer to SCID mice. Specific monoclonal human Fab fragments can then be isolated from the mice by repertoire cloning even when the human donor's last contact with antigen was more than 17 years ago.

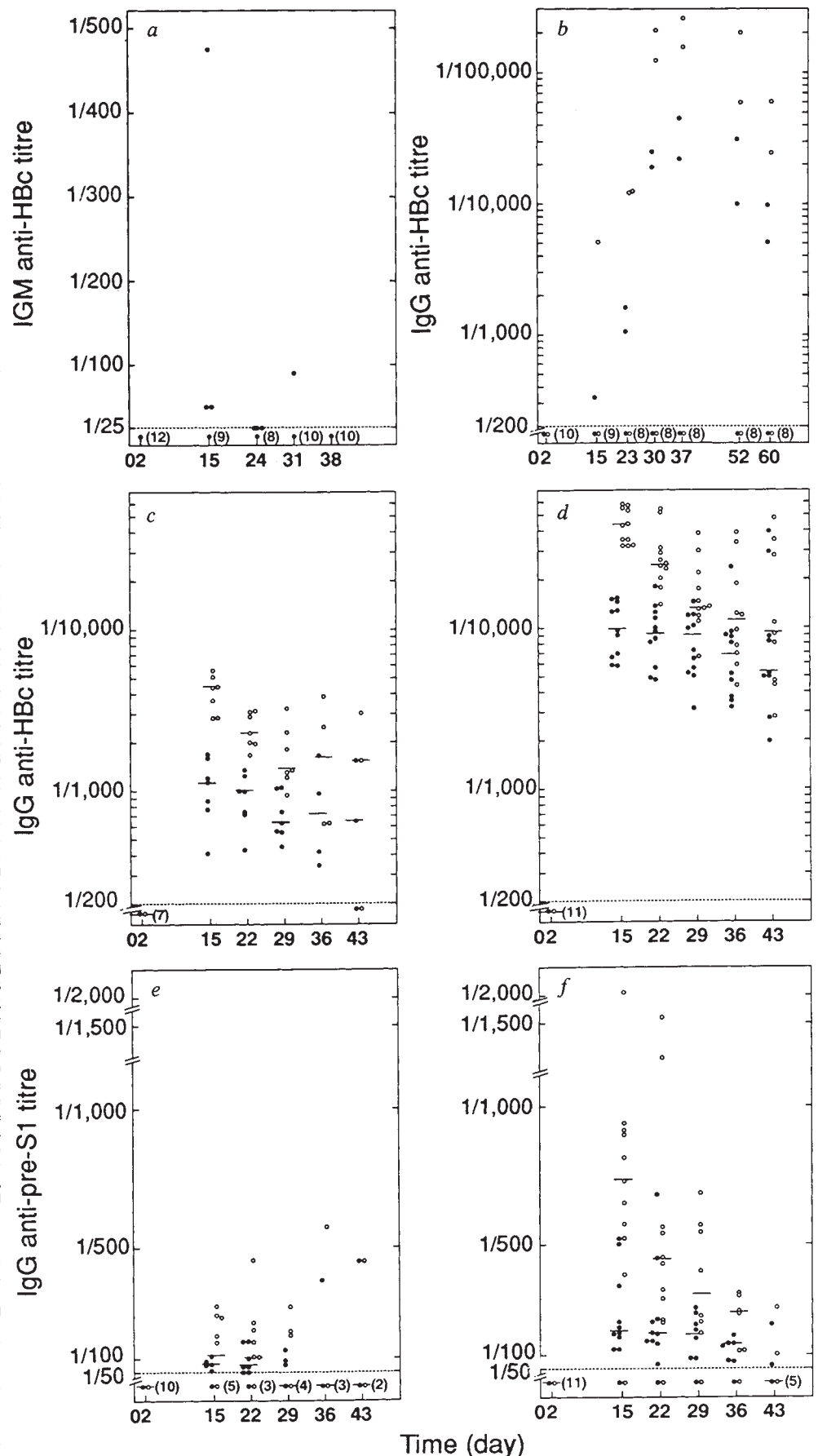
The degree of functionality of the human cell xenograft in SCID mice, notably the ability to produce a specific immune response, has been controversial and potentially limiting^{9,10,12-14}. Using our refined peripheral blood lymphocyte (PBL) transfer protocol (M.A.D., S.A.E., P.J.M. and F.J.D., manuscript submitted), we have now established procedures for immunization of hu-PBL-SCID mice (see Fig. 1 legend). Human PBL are generally exposed to antigen *in vitro* before injection into mice and again *in vivo* after cell transfer. Several mice are used in each experiment.

Hepatitis B core (HBc) antigen¹⁵ was chosen for initial studies. Using PBL from a donor (number 1) with no detectable IgG anti-HBc serum titre, all nonimmunized hu-PBL-SCID mice were found to lack such antibody. A small proportion of mice produced a weak transient IgM response after HBc immunization (Fig. 1a). When PBL from a donor (number 2) with a low IgG anti-HBc serum titre were used, all nonimmunized hu-PBL-SCID mice again had no specific antibody titre. After

FIG. 1. Immunization with HbC and pre-S1 antigens. **a**, IgM anti-HbC responses in 12 hu-PBL-SCID mice populated with cells derived from donor number 1 (no detectable IgG anti-HbC titre and no history of hepatitis) and immunized with HbC. No IgG anti-HbC titre was detected. Twelve similarly populated, non-immunized, hu-PBL-SCID mice lacked both IgM and IgG anti-HbC titres. **b**, IgG anti-HbC responses in 10 mice derived from donor number 2 (low IgG anti-HbC titre: 1/320); and **d**, in 11 mice derived from donor number 3 (intermediate IgG anti-HbC titre: 1/7,700) immunized with HbC. Sixteen nonimmunized hu-PBL-SCID mice derived from donor number 2 had no detectable human IgG anti-HbC titres. **c**, Seven of seven tested nonimmunized hu-PBL-SCID mice derived from donor number 3 had low, but detectable IgG anti-HbC. **f**, IgG anti-pre-S1 titres in 11 mice derived from donor number 3 (with low IgG anti-pre-S1 serum titre: 1/185) and immunized with HbC-pre-S1, and **e**, in 11 nonimmunized controls. Each dot corresponds to an individual bleed, unless indicated by a number in parenthesis. Whenever appropriate, median values are indicated by horizontal bars. Some hu-PBL-SCID mice died during the experiment, resulting in the smaller number of sera examined with time. ●, Absolute titre; ○, normalized titre. All the donors in this study have anti-Epstein-Barr virus antibody detected as described²³.

METHODS. Lymphophereses were taken from consenting donors. Nonleaky²⁴ C.B-17 SCID (SCID) mice were populated with PBL before they were 9 weeks old. For immunization of hu-PBL-SCID mice, 50×10^6 PBL were incubated (diluted to a density of $4 \times 10^6 \text{ ml}^{-1}$ using RPMI/10% FCS) for 4 h with $2 \mu\text{g ml}^{-1}$ purified recombinant HbC or HbC-pre-S1 in a humidified incubator at 37°C with 7% CO_2 . The cells were then washed once in Earle's (Flow) isotonic medium and injected in 1 ml of Earle's medium. These mice were then injected subcutaneously with $10 \mu\text{g}$ HbC or HbC-pre-S1 in complete Freund's adjuvant at the base of the tail on day 2, and with the same dose of antigen subcutaneously in saline on day 22 (**d**, **f**), 23 (**b**) or 24 (**a**). Control mice were processed in parallel without antigen. hu-PBL-SCID mice were bled sequentially, and human IgG serum levels determined as described²⁵. IgM- and IgG-anti-HbC, and IgG anti-pre-S1 titres were determined by indirect ELISA. Microtitre plates (Costar) were coated with HbC, or with a mixture of peptides corresponding to the adw2, and ayw subtypes of the HBV amino-acid sequence 12–47 (pre-S1)^{16,26–28}. Nonspecific binding was prevented by incubation with PBS/5% nonfat dairy milk. Serial sera dilutions from hu-PBL-SCID mice and donors were further incubated in duplicate. The specific antibodies were revealed using peroxidase-labelled, affinity-purified, goat anti-human IgM or IgG (Cappel), followed by the addition of *O*-phenylenediamine (Sigma). The reactions were stopped by the addition of H_2SO_4 , and the optical density determined at 492 nm. After each step, several washes with PBS 0.05% Tween 20 were done. The titre corresponding to an optical density of twice that of normal negative donor serum at lowest dilution was recorded (absolute value). The human specificity of these assays were confirmed by their nonreactivity with sera from unmanipulated SCID, C57BL/6 and BALB/c mice nonimmunized and immunized with HbC. The three hu-PBL-SCID mice positive for IgM anti-HbC were not leaky²⁴ at the time of testing. As the levels of serum human immunoglobulin are variable between analysed mice, and increased dramatically in a single SCID mouse recipient after PBL transfer, we also derived normalized values for IgG calculated as follows: absolute titre $\times$ (IgG serum level (donor)/IgG serum level (hu-PBL-SCID mouse sample)). This allowed unambiguous differentiation between passive and active changes of the antibody tested, and therefore a reliable assessment of the putative immune response in a given mouse with time or between groups of hu-PBL-SCID mice derived from a common donor. Normalized levels between hu-PBL-SCID mice derived from donors having similar IgG serum levels (see Fig. 2) can also be compared with each other. Both pFS14 and HbC-pre-S1 plasmids are composed of the full-length sequence of the gene for the HbC protein, the HbC-pre-S1 plasmid contains an additional upstream sequence corresponding to the pre-S1 amino-acid region 12–47 (refs 16, 26–28, and Ann M. Moriarty and G.B.T., manuscript in preparation). HbC and HbC-pre-S1 antigens were prepared from lysis²⁹ of IPTG-induced *E. coli* carrying the respective expression vector. After DNaseI digestion, the proteins were

further purified by selective centrifugation, gel filtration (Bio-Gel-A-50m, Biorad), and finally hydroxylapatite chromatography (Bio-Gel-HT, Biorad).



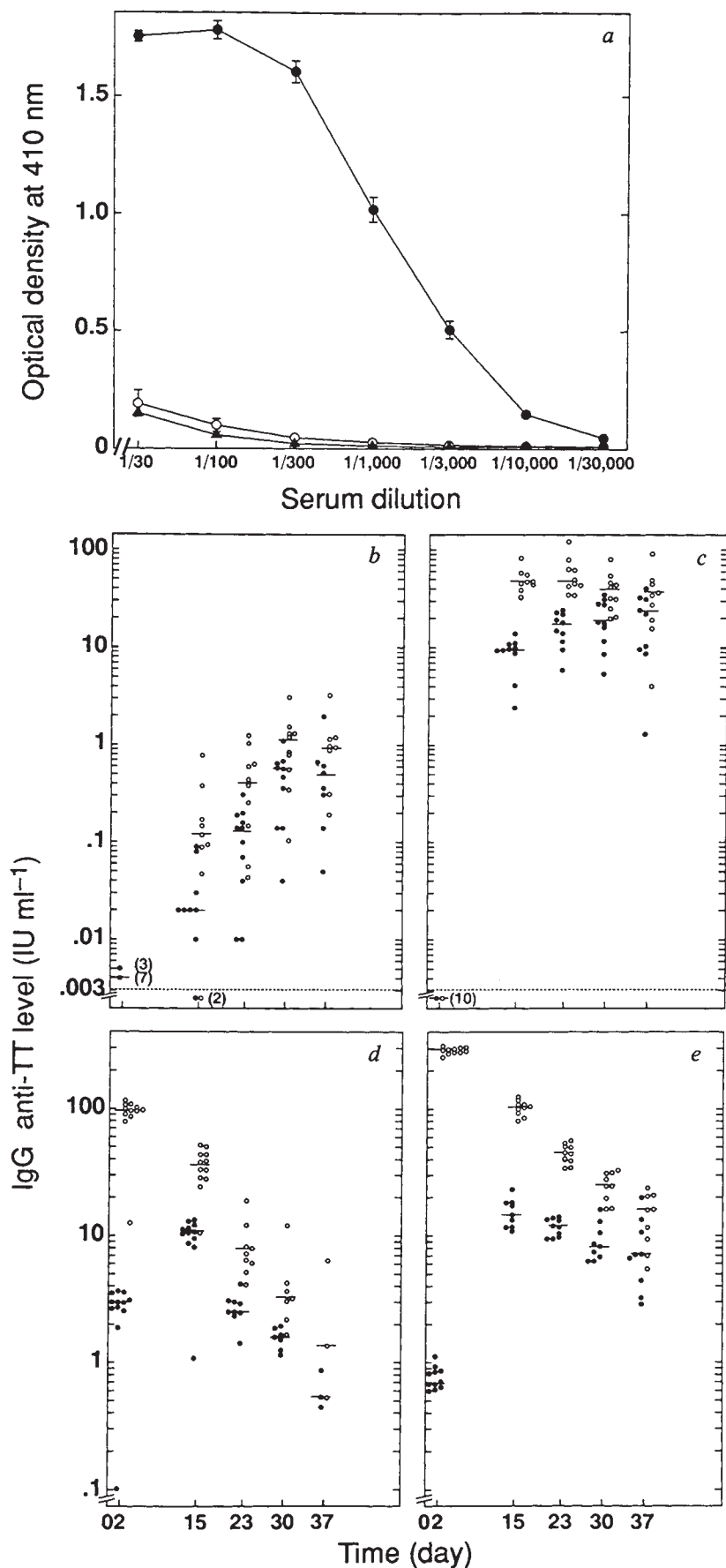


FIG. 2 Immunization with TT antigen. *a*, Comparison of TT serum titre for a donor (number 4) not boosted for 20 years (▲) with that ($\pm$ s.e.m.) obtained by boosting the donor's PBL in six hu-PBL-SCID mice (●). These six hu-PBL-SCID mice had, 29 days after cell transfer, 2.86 ± 1.19 and 6.36 ± 3.34 IU of mean absolute, and normalized IgG anti-TT ml⁻¹, respectively. The mean values for three similarly populated mice but not immunized is represented (○). This ELISA did not show any reactivity when sera from nonpopulated SCID, BALB/c or C57BL/6 mice were applied. IgG anti-TT serum responses in hu-PBL-SCID mice nonimmunized (*b*, *d*) or immunized (*c*, *e*) with TT. Mice were populated with PBL from either donor number 5 (*b*, *c*), immune to TT with an IgG anti-TT level of 0.88 IU ml⁻¹, not boosted with TT for at least 4 years, or from donor number 6 (*d*, *e*), immune to TT, and boosted with TT (Wyeth) 6 days before PBL transfer. IgG anti-TT serum levels of donor number 6 at day 0, and day 37 were 0.65 and 7.60 IU ml⁻¹, respectively. Donors number 4, 5 and 6 have IgG serum levels of 8.49, 8.12 and 9.82 g l⁻¹, enabling comparisons between groups of hu-PBL-SCID mice normalized anti-TT levels. Symbols as in Fig. 1.

METHODS. Immunized and control groups of hu-PBL-SCID mice were handled as in Fig. 1 with the following modification: in the immunized group, PBL were incubated *in vitro* with 15 μ g ml⁻¹ TT antigen (Wellcome), and mice injected with 50 μ g TT on day 2, and on day 9 (*a*) or 23 (*c*, *e*). IgG anti-TT levels were determined as previously described, and the threshold of the assay was 0.003 IU ml⁻¹ (M.A.D. *et al.*, manuscript submitted).

immunization, a small proportion of the mice showed strong IgG anti-HBc responses with serum antibody levels 10^2 – 10^3 -fold greater than the donor (Fig. 1b). This individual variation probably arises from variation in the cells received by each mouse. All of the mice with cells derived from a donor (number 3) with a higher IgG anti-HBc titre (1/7,700), but not immunized, developed specific IgG serum titres (Fig. 1c) at levels below that of the donor. Mice similarly populated responded to immunization with median IgG anti-HBc titres over 10-fold higher than those found in the nonimmunized mice (Fig. 1d). This HBc immunization had no effect on the IgG anti-tetanus-toxoid serum levels of these hu-PBL-SCID mice (data not shown). One donor (number 3) carried low levels of antibody to the pre-S1 antigen¹⁶ of hepatitis B virus (HBV; Fig. 1 legend), which is less immunogenic than HBc. We immunized hu-PBL-SCID mice derived from this donor with a recombinant protein in which a sequence from the pre-S1 region was linked to, and so could benefit from, the carrier effect of the HBc protein¹⁷ (HBc-pre-S1 antigen). A brisk IgG anti-HBc response was seen in all immunized mice (data not shown), and these mice had peak median normalized IgG anti-pre-S1 titres roughly fourfold higher than the donor, with one mouse having an 11-fold increase. These titres were much higher than those from non-immunized mice (Fig. 1e,f).

Tetanus toxoid (TT) was chosen for further studies because the date of last contact of the donor with antigen could be ascertained. A number of donors were used. One donor (number 4) had not been boosted with TT for at least 20 years and had no significant serum IgG anti-TT (Fig. 2a). Production of anti-TT antibody was restored to high levels in all of the immunized mice but was virtually undetectable in the non-immunized animals (Fig. 2a). Using another donor (number 5) having a significant serum level of IgG anti-TT, but not boosted for at least four years, all nonimmunized mice had detectable IgG anti-TT levels (Fig. 2b). After immunization, median normalized IgG anti-TT levels in the mice showed a 55-fold increase relative to the donor (in an individual up to a 140-fold increase), and a 50–450-fold increase relative to the nonimmunized group (Fig. 2c). Using a donor (number 6) recently boosted with TT, the increases in IgG anti-TT levels in immunized versus nonimmunized mice were less marked than with nonboosted donors (Fig. 2d,e). The anti-TT response in hu-PBL-SCID mice was, as in humans¹⁸, generally highly skewed toward IgG1 (M.A.D. *et al.*, manuscript in preparation).

For repertoire cloning experiments, a donor (number 1) was used who had significant serum levels of IgG anti-TT (1.99 IU ml^{-1}) but had not been boosted with antigen for 17 years. RNA was extracted from blood and tissues of a hu-PBL-SCID mouse which had produced a large increase in serum IgG anti-TT two weeks after the last *in vivo* boost (37 days after PBL transfer, absolute and normalized IgG anti-TT levels of 5.0 and 66.8 IU ml^{-1} , respectively). Polymerase chain reaction (PCR) amplification using γ_1 and κ chain-specific primers³ showed the presence of human antibody messenger RNA in spleen, peripheral blood, kidney and liver. The spleen RNA was used to construct Fab (γ_1 , κ) combinatorial libraries in phage λ (refs 3, 19). Screening of the library revealed two positive clones in 370,000. They were plaque-purified and the phagemids were excised and used to transform *Escherichia coli*. Fabs from these transformants were subjected to competition enzyme-linked immunosorbent assay (ELISA) to show apparent binding affinities (that is, reciprocal 50%-inhibition concentrations) of 10^8 – 10^9 M^{-1} (Fig. 3a) comparable to Fabs isolated directly from a boosted individual³. Nucleic-acid sequencing revealed that the clones were different from one another (Fig. 3b).³

Screening of an Fab library of one million clones from a nonboosted individual with a serum level of IgG anti-TT similar to donor number 1 failed to reveal any positives³. This is expected, given the low resting level of plasma cells (and specific mRNA) in peripheral blood^{6,20}.

From the same spleen RNA an Fab (γ_1 , λ) library was constructed in the M13 phage surface expression vector pComb3 (ref. 21). This was panned against TT when a large number of positives were identified and 10 clones were sequenced. The heavy chains fell into two groups, one of which had been seen in the phage λ library, and all the λ light chains appeared related (illustrated in Fig. 3b for four clones). Therefore PBL from individuals who have not had contact with antigen for many years can be stimulated to produce specific antibody using this

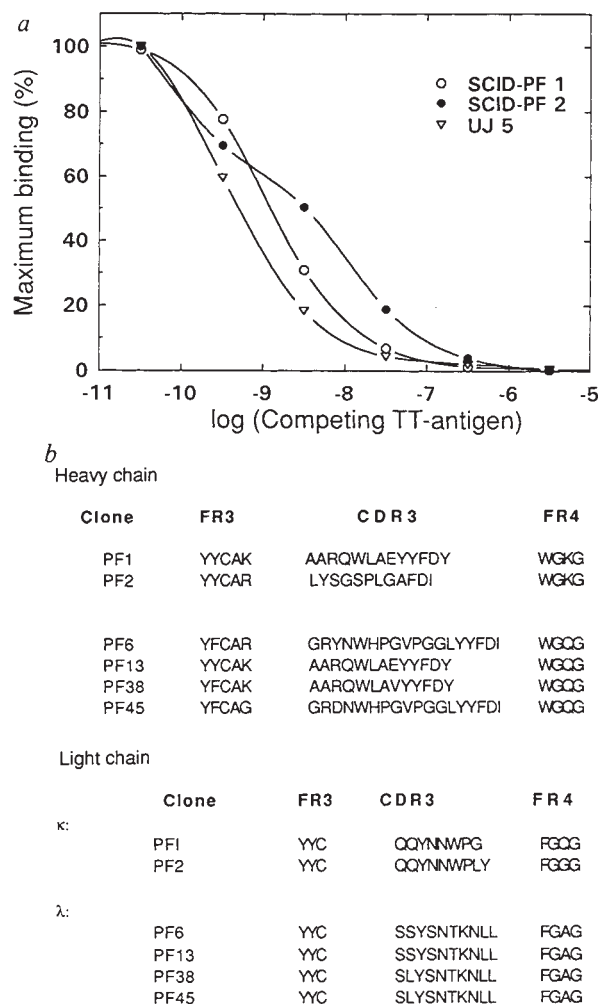


FIG. 3 Characterization of TT-binding Fabs selected from phage combinatorial libraries prepared from a TT-boosted in hu-PBL-SCID mouse derived from donor number 1. **a**, Specificity of antigen binding shown by competitive ELISA. ○, ●, Fabs isolated from a hu-PBL-SCID mouse combinatorial library in phage lambda. ▽, Fab isolated from a human combinatorial library³. **b**, Sequences of the variable heavy (V_H) and variable light (V_L) chain CDR3 regions of the anti-TT Fabs (single-letter amino-acid code). Interestingly, the PF1 and PF13 clones have heavy-chain CDR3 sequences identical to that of a previously described TT-binder¹³, but there are several differences elsewhere in the sequences. The clones PF1 and PF2 (IgG1 κ) were derived from a library in phage λ (ref. 19) and the other clones (IgG1 λ) from a library in the M13 phage surface expression vector pComb3 (ref. 21).

METHODS. The combinatorial library in phage λ was constructed as previously reported³, but with the addition of three new oligonucleotide primers for the DNA amplification of the heavy γ_1 , and κ chain, respectively (V_H 2f=5'-CAGGTGCAGTACTCGAGTCGGG-3'; V_H 4f=5'-CAGGTGCAGTCTCGAGTCGGG-3'; V_K 2a=5'-GATATTGAGTCACTCACTCTCCA-3'). The competitive ELISA assay, and nucleic-acid sequencing were done as described³. The combinatorial library in the M13 phage surface expression vector pComb3 was constructed as described²¹. The heavy-chain PCR primers were as above and the lambda light chain primers as in ref. 30. The library (3×10^6 members) was taken through three rounds of panning when 43 out of 46 clones reacted with TT in a colony-screening assay²¹.

SCID mouse model. These antibodies can then be cloned as Fab fragments using the combinatorial approach. This has evident consequences. Among them, human autoantibodies, potentially useful for immunosuppressive and immunomodulatory function, could be obtained using appropriate autoimmune donors in the model. Similarly, antibodies against specific MHC

determinants for treating graft rejection could be derived from selected multiparous women²². Expanded repertoires of anti-HIV antibodies against specific epitopes could be generated by peptide stimulation of PBL from seropositive donors. Moreover this model could serve to study the factors influencing secondary (auto)immune responses *in vivo*. □

Received 30 May; accepted 17 October 1991.

1. Winter, G. & Milstein, C. *Nature* **349**, 293–299 (1991).
2. James, K. & Bell, G. T. *J. Immunol. Meth.* **100**, 5–40 (1987).
3. Persson, M. A. A., Caothien, R. H. & Burton, D. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2432–2436 (1991).
4. Caton, A. J. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6450–6454 (1990).
5. Caton, A. J. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1590 (1991).
6. Burton, D. R. *Trends Biotechnol.* **9**, 169–175 (1991).
7. Lerner, R. A., Barbas, C. F. III, Kang, A. S. & Burton, D. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9705 (1991).
8. Bosma, G. C. Custer, R. P. & Bosma, M. J. *Nature* **301**, 527–530 (1983).
9. McCune, J. M. *et al. Science* **241**, 1632–1639 (1988).
10. Mosier, D. E., Gulizia, R. J., Baird, S. M. & Wilson, D. B. *Nature* **335**, 256–259 (1988).
11. Mosier, D. E., Gulizia, R. J., Baird, S. M. & Wilson, D. B. *Nature* **338**, 211 (1989).
12. Bankert, R. B. *et al. Curr. Topics Microbiol. Immun.* **152**, 201–210 (1989).
13. Tighe, H. *et al. Eur. J. Immun.* **20**, 1843–1848 (1990).
14. Krams, S. M., Dorshkind, K. & Gershwin, M. E. *J. exp. Med.* **170**, 1919–1930 (1989).
15. Milich, D. R. & McLachlan, A. *Science* **234**, 1398–1401 (1986).
16. Tiollais, P., Pourcel, C. & Dejean, A. *Nature* **317**, 489–495 (1985).
17. Milich, D. R., McLachlan, A., Thornton, G. B. & Hughes, J. L. *Nature* **329**, 547–549 (1987).
18. Seppälä, I. J. T., Routonen, N., Sarnesto, A., Mattila, P. A. & Mäkelä, O. *Eur. J. Immun.* **14**, 868–875 (1984).
19. Huse, W. D. *et al. Science* **246**, 1275–1281 (1989).
20. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. *Nature* **352**, 624–628 (1991).

21. Barbas, C. F. III, Kang, A. S., Lerner, R. A. & Berkovic, S. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7978–7982 (1991).
22. Hansen, T. H. & Sachs, D. H. in *Fundamental Immunology* (ed. Paul, W. E.) 445–487 (Raven, New York, 1989).
23. Cannon, M. J., Pisa, P., Fox, R. I. & Cooper, N. R. *J. clin. Invest.* **85**, 1333–1337 (1990).
24. Bosma, G. C. *et al. J. exp. Med.* **167**, 1016–1033 (1988).
25. Duchosal, M. A., McConahey, P. J., Robinson, C. A. & Dixon, F. J. *J. exp. Med.* **172**, 985–988 (1990).
26. Valenzuela, P. *et al. Nature* **280**, 815–819 (1979).
27. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. *Nature* **281**, 646–650 (1979).
28. Pasek, M. *et al. Nature* **282**, 575–579 (1979).
29. Stahl, S., MacKay, P., Magazin, M., Bruce, S. A. & Murray, K. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1606–1610 (1982).
30. Williamson, R. A., Persson, M. A. A. & Burton, D. R. *Biochem. J.* **277**, 561–563 (1991).

ACKNOWLEDGEMENTS. We thank R. A. Lerner for support and encouragement; R. A. Reisfeld, M. A. A. Persson, E. Golub and B. Toyonaga for discussions; A. M. Moriarty for the HBC-pre-S1 plasmid, F. Schödel for the pS14 plasmid, R. S. Balderas for subclass specific antihuman IgG antibodies, J. P. Dickson and D. Hoekstra for assistance with sequencing; H. Fox and M. Eddleston for critically reading the manuscript and M. K. Occhipinti for editing it. M.A.D. is the recipient of a fellowship from the Terry Gottlieb Lupus Research Institute. P.F. is a Liebig Fellow of the Stiftung Stipendien-Fonds des Verbandes der Chemischen Industrie e.V. D.R.B. is a Jenner Fellow of the Lister Institute of Preventive Medicine. This work was supported by the NIH and Johnson & Johnson.

A chimaeric 11 β -hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension

Richard P. Lifton^{*†}, Robert G. Dluhy[†], Michael Powers^{*}, Glenn M. Rich[†], Sandra Cook[†], Stanley Ulick[‡] & Jean-Marc Lalouel^{*}

^{*} Howard Hughes Medical Institute, University of Utah, Eccles Institute of Human Genetics, Suite 6200, Salt Lake City, Utah 84132, USA

[†] Endocrine-Hypertension Division, Brigham and Women's Hospital and Harvard Medical School, 221 Longwood Avenue, Boston, Massachusetts 02115, USA

[‡] Veterans Affairs Hospital, Bronx, New York 10468, USA

GLUCOCORTICOID-REMEDIALD aldosteronism (GRA), an autosomal dominant disorder, is characterized by hypertension with variable hyperaldosteronism^{1,2} and by high levels of the abnormal adrenal steroids 18-oxocortisol and 18-hydroxycortisol^{3–5}, which are all under control of adrenocorticotrophic hormone and suppressible by glucocorticoids⁶. These abnormalities could result from ectopic expression of aldosterone synthase, which is normally expressed only in adrenal glomerulosa, in the adrenal fasciculata. Genes encoding aldosterone synthase^{7–9} and steroid 11 β -hydroxylase⁷ (expressed in both adrenal fasciculata and glomerulosa), which are 95% identical⁷ and lie on chromosome 8q (refs 7, 10), are therefore candidate genes for GRA. Here we demonstrate complete linkage of GRA in a large kindred to a gene duplication arising from unequal crossing over, fusing the 5' regulatory region of 11 β -hydroxylase to the coding sequences of aldosterone synthase (maximum lod score 5.23 for complete linkage, odds ratio of 170,000:1). This mutation can account for all the physiological abnormalities of GRA. Our result represents the demonstration of a mutation causing hypertension in otherwise phenotypically normal animals or humans.

Figure 1 shows GRA kindred 2061, clinically characterized by early hypertension and death from stroke before age 45

(G.M.R. *et al.*, manuscript submitted). Hybridization of exons 3–4 of 11 β -hydroxylase (11-OHase) to *Bam*HI-digested DNA reveals a novel 6.3-kilobase(kb) fragment present only in affected pedigree members (Fig. 2a) and absent in 90 unrelated individuals. This anomalous fragment yields a maximum lod score of 5.23 for complete linkage with GRA. The intensity of this fragment is 50% of that observed for either the normal aldosterone synthase (AldoS) or 11-OHase fragments (Fig. 2b), compatible with the presence of one additional copy of an AldoS/11-OHase-like gene in affected subjects, but inconsistent with a simple site polymorphism or insertion/deletion.

A duplication at this locus could potentially arise by unequal crossing over between homologous regions of 11-OHase and AldoS (Fig. 3a). Fusion of 5' regulatory sequences of 11-OHase to a coding region with aldosterone synthase activity could lead to the physiological abnormalities of GRA. This chimaera could only result if the two genes were closely linked in the orientation shown in Fig. 3a.

This hypothesis predicts specific structural features of the extra AldoS/11-OHase gene in this pedigree. First, only unequal crossing-over occurring between the *Bam*HI sites 5' of exon 1 and within intron 5 of AldoS would generate an anomalous fragment detectable by the exon 3–4 probe (Fig. 3b). The size of the anomalous fragment produced by this crossover would be 6.3 kb, which is the size of the observed anomalous fragment. Second, the sequences contained on this fragment should extend from the *Bam*HI site in the 5' flanking region of 11-OHase to the *Bam*HI site unique to intron 5 of AldoS. This structure is confirmed by separate hybridization to *Bam*HI-cut DNA of exons 1–9 and 5' flanking sequences unique to either 11-OHase or AldoS (Fig. 2c). Third, unequal crossing over should result in characteristic fragments upon digestion of genomic DNA with various restriction endonucleases (Fig. 3b). In each case digestion produces the novel fragments predicted or, equally importantly, fragments indistinguishable from those derived from 11-OHase and AldoS (Fig. 2d and 4a). Fourth, AldoS sequences 3' to the crossover should be duplicated, but 5' AldoS sequences and 3' 11-OHase sequences should not. We have identified specific AldoS and 11-OHase sequences that are polymorphic by single-strand conformational polymorphism¹¹. In heterozygotes, the number of gene copies on affected and unaffected chromosomes can be compared by quantitation¹². With AldoX9, an exon-9 AldoS-specific marker, all unaffected

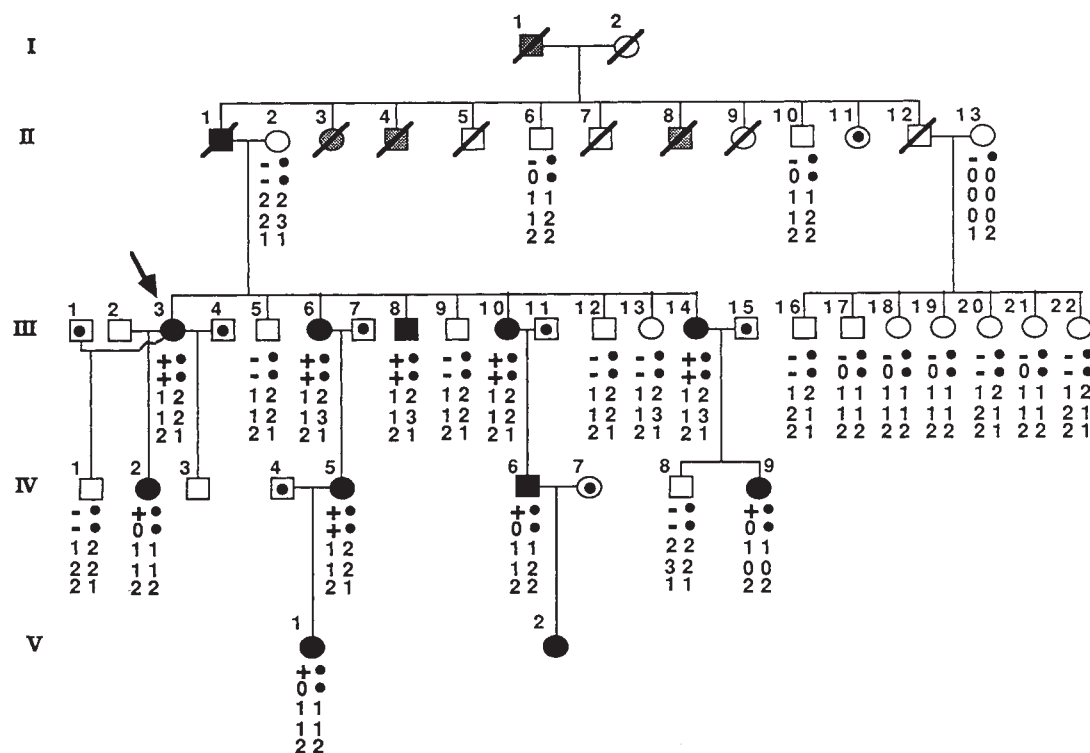
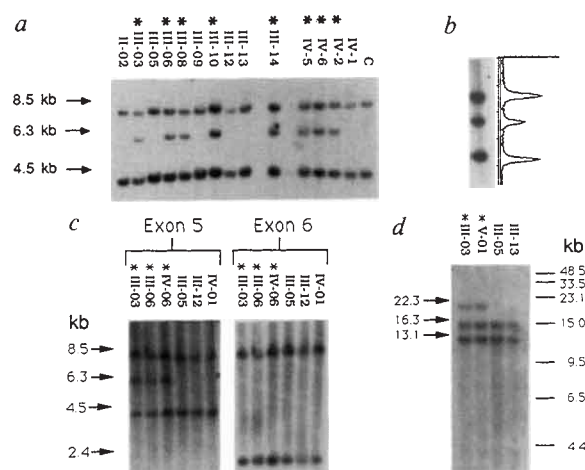


FIG. 1 Kindred 2061. Living individuals with GRA (black symbols), without GRA (open symbols), or unevaluated (black dots) are shown. All deceased individuals except II-1 were classified as affection unknown, although clinical history of death from stroke before age 45 strongly suggests presence of GRA in some (grey symbols). II-1 had severe hypertension, died from stroke at age 42 and had 5 affected offspring by an unaffected spouse. Genotypes at 5 marker loci are shown. In descending order these are: presence (+) or absence (−) of the 6.3-kb *Bam*HI fragment hybridizing to exon 3–4 of 11-OHase; presence (+) or absence (−) of extra copy of allele 1 of exon 9 of AldoS revealed by marker AldoX9, scorable only in heterozygotes for AldoX9 (Fig. 4b); alleles detected by markers AldoX9, specific for exon 9 of AldoS; AldoX1, specific for the 5' flanking region of AldoS; 11OH-3, specific for intron 5 of 11-OHase. Affection status was assessed by measurement of urinary 18-oxotetrahydrocortisol¹⁸, and the ratio of this metabolite to urinary tetrahydroaldosterone; individuals designated as affected all showed values diagnostic of GRA⁶. All affected individuals were either frankly hypertensive or had age- and sex-adjusted blood pressure greater than

90th percentile (G.M.R. *et al.*, manuscript submitted). In linkage analysis the gene frequency of GRA was specified at 0.0001 with autosomal dominant inheritance, sporadic rate of 0.00005 and penetrance of 95%. Changes in parameters or affection status of II-1 had small effects on the lod score. METHODS. Polymerase chain reaction (PCR)-based markers were genotyped by amplification of genomic DNA in the presence of ³²P-labelled dCTP; unless specified, the PCR was for 30 cycles of 94 °C for 45 s, 60 °C for 45 s, 72 °C for 45 s. Products were analysed by single-strand conformational polymorphism¹¹. Primers were designed from published sequences⁷; marker 11OH-3: primers RL11-3 (5'-TTACTTGGGATTGTGATGTGATAAAC-3') and RL11-4 (5'-CACCGTCTGCAGGAGACAG-3'). Marker AldoX9: primers RLAlldoX9 (5'-CTGGGCTTGGCCAGGCTTGTG-3') and RLAlldoX10 (5'-GACAGGGACATGTGAGACTAG-3'). Marker AldoX1: primers RLAlldoX1 (CCCACAGCATGTGACCACCAG) and RLAlldoX2 (AGAAAAGGCGTGGGTCTGGAC). Variants detected with these markers were shown to be allelic by inheritance, and specificity of markers was demonstrated by PCR with human cosmid clones.

FIG. 2 a, Hybridization of exons 3–4 of 11-OHase to *Bam*HI-digested genomic DNA of portion of K2061. Affected subjects are indicated by asterisks; C represents a control. Positions are shown of the 8.5-kb fragment from 11-OHase, the 4.5-kb fragment from AldoS and the aberrant 6.3-kb fragment. b, Densitometry of fragment bands. DNA from a lymphoblastoid cell line of individual III-10 was prepared and analysed as in a. Densitometric traces are shown and quantitation with a Molecular Dynamics Phosphorimager reveals a 2:1:2 ratio of the 8.5-, 6.3- and 4.5-kb fragments, respectively. c, Hybridization of exon 5- and exon 6-specific sequences to *Bam*HI-digested DNA of affected and unaffected subjects. Both probes hybridize to the expected 11-OHase and AldoS fragments. In addition, the exon 5 probe hybridizes to the 6.3-kb anomalous *Bam*HI fragment in affected subjects whereas the exon 6 probe does not. d, Hybridization of exon 3–4 of 11-OHase to *Eco*RI-digested DNA of affected and unaffected subjects of K2061. DNA was fractionated on a 0.5% agarose gel then hybridized to exons 3–4. The positions of phage lambda size standards (in kb) are indicated at the right of the figure. All subjects display 16.3- and 13.1-kb fragments derived from the normal 11-OHase and AldoS genes, respectively. In addition, the two affected subjects display a 22-kb fragment not present in the two unaffected subjects or in 90 unrelated controls. The sizes of the fragments derived from 11-OHase and AldoS measured here are somewhat smaller than previously reported⁷.

METHODS. Genomic DNA preparation¹⁹, Southern blotting and hybridization²⁰ were as previously described. Hybridization probes were labelled either by random priming of isolated fragments²¹ or by PCR²² using appropriate primers and template c110H26, a human cosmid clone containing the



11-OHase gene (R.P.L., unpublished results). Primers for exons 3–4 were: RL11E3-1 (5'-CAAGCTCTGCCCTGGCCTCTG-3') and RL11E4-2 (5'-GGTGTCTTCCCATAGCACTG-3'). Primers for exon 5: RL11E5-1 (5'-CACTGAAGGATGCTTCCAGCAC-3'), and RL11E5-2 (5'-CTCTCTGGTGGGCTGGTTG-3'). Primers for exon 6: RL11E6-1 (5'-CTGTGTCTCTGCAGACGGTG-3') and RL11E6-2 (5'-CAGGGCCACAGGGAGGCCTC-3').

a

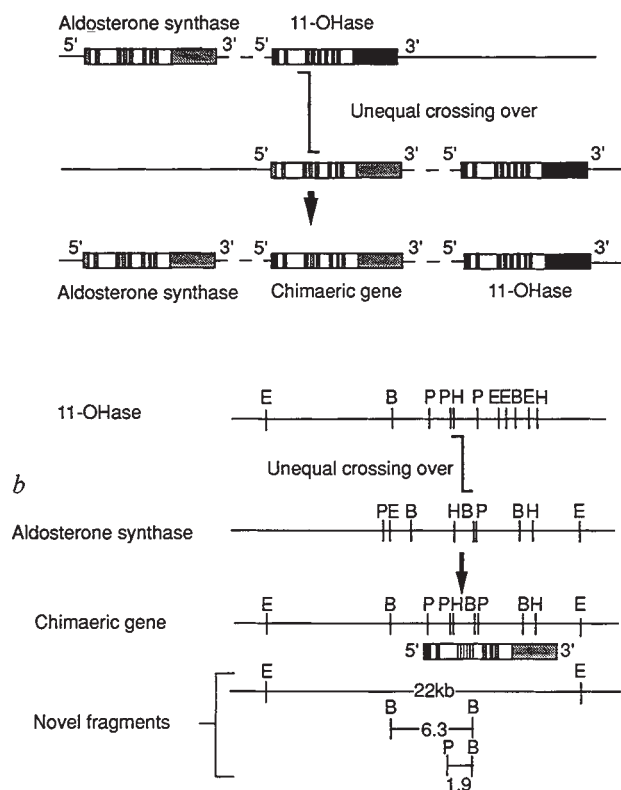


FIG. 3 a, Diagram depicting unequal crossing over between aldosterone synthase and 11-OHase genes. Each gene is represented by a wide bar, with the location of exons⁷ indicated by either black (11-OHase) or stippled (aldosterone synthase) bands. One of the two products of unequal crossing over will have a chimaeric gene located between normal AldoS and 11-OHase genes. In this example unequal crossing over is depicted as occurring in the intron between exons 3 and 4. The true physical distance between 11-OHase and AldoS is unknown. b, Location of cleavage sites for restriction endonucleases in 11-OHase, aldosterone synthase and the hypothesized chimaeric gene resulting from unequal crossing over. Sites cut by *Bam*HI (B), *Eco*RI (E), *Hind*III (H), and *Pvu*II (P) are shown and the location of the exons of the chimaeric gene on the restriction enzyme map are indicated. The map predicts that digests of DNA of affected subjects with the enzymes indicated either alone or in combination with *Bam*HI should together give 3 novel fragments hybridizing to exon 3-4. All other products of these digests should be identical in size to fragments derived from either 11-OHase or AldoS. The origin of each of these novel fragments is demonstrated below the map of the chimaeric gene. The map of sites in 11-OHase and AldoS is as before⁷, with the exception of a *Pvu*II site located in intron II of 11-OHase, a region not sequenced previously. A cluster of 4 *Pvu*II sites and a *Bam*HI site common to the 3' ends of all 3 genes are not shown.

heterozygotes show 1:1 amplification of alleles 1 and 2. In contrast, a 2:1 ratio of these alleles is seen in all affected heterozygotes, indicating the presence of an extra copy of the AldoS sequences (Fig. 4b). Linkage analysis reveals maximum lod scores of 3.15 and 3.25 for complete linkage of this extra copy to GRA and the duplication detected on Southern blotting, respectively. But markers unique to the 5' flanking region of AldoS (AldoX1) and intron 5 of 11-OHase (110H-3) are not duplicated in affected subjects (data not shown).

Finally, unequal crossing over requires tight linkage of native 11-OHase and AldoS genes. Markers AldoX1/AldoX9 and 110H-3 show a maximum lod score of 3.63 for complete linkage in K2061. Furthermore, no recombinants are detected between the duplication and the native 11-OHase/AldoS genes (lod score is 1.63 for complete linkage).

These results demonstrate that in K2061, GRA is caused by a mutation tightly linked to a chimaeric 5' 11-OHase/AldoS 3' gene. The structure of this gene indicates an origin from unequal crossing over upstream of intron 5. This chimaeric gene is the likely cause of GRA in this pedigree. We propose that the product of this gene has aldosterone synthase activity due to

coding sequences of AldoS, which are aberrantly expressed in adrenal fasciculata because of 11-OHase regulatory sequences. This would result in (1) activity of aldosterone synthase on cortisol and cortisol precursors, resulting in synthesis of 18-oxocortisol, 18-hydroxycortisol and aldosterone in adrenal fasciculata; (2) control of synthesis of these steroids by adrenocorticotrophic hormone (ACTH) (as 11-OHase regulatory sequences are under the control of ACTH in adrenal fasciculata¹³) and consequent glucocorticoid suppressibility; (3) a neomorphic mutation with autosomal dominant inheritance. These features are the hallmarks of GRA¹⁻⁶. Further studies will pinpoint the site of unequal crossing over, demonstrate activity of the gene product, and determine whether other GRA patients harbour similar mutations.

GRA is considered a rare disease, although systematic screening of hypertensive subjects has not been performed. That GRA can result from unequal crossing over raises the possibility that such mutations may be more common than is presently recognized. Moreover, different sites of unequal crossing over might result in varying phenotypic effects. This mutation is reminiscent of neomorphic haemoglobinopathies and 21-hydroxylase

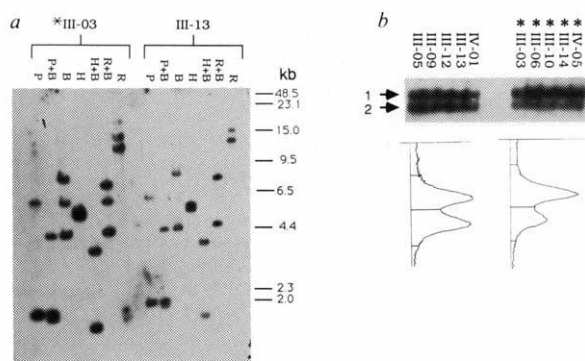


FIG. 4 a, Hybridization of exon 3-4 probe to genomic DNA of an affected and an unaffected subject digested with various endonucleases (see legend to Fig. 3b for single-letter nuclease designation; R is *Eco*RI). The products of all digests are those predicted from unequal crossing over (Fig. 3b). Results in another pair of affected and unaffected subjects are identical (not shown). b, Single-strand conformational polymorphism of marker AldoX9 in 5 unaffected subjects (left) and 5 affected subjects (right). Autoradiogram (top) and densitometry (bottom) are shown. Relative intensity of allele 1:allele 2 (arrowed bands) is 1:1 in unaffected subjects and 2:1 in affected subjects. Results were identical for 3 independent amplifications, and were unaffected by changing annealing temperature, number of cycles, or concentration of genomic DNA in PCR. Quantitation of duplicate samples reveals, in affected individuals, a mean ratio of allele 1 to allele 2 of 1.95:1 (standard deviation, 0.07; range, 1.85-2.04); in unaffected individuals the mean allelic ratio in 1.05 (standard deviation, 0.05; range, 1.01-1.12).

deficiency, in which unequal crossing over generates null alleles¹⁴.

Although the genetics of hypertension^{15,16} generally seem to favour polygenic inheritance¹⁷, major gene effects in subsets of the hypertensive population have not been excluded. We have demonstrated the feasibility of identifying mutations causing mendelian forms of hypertension. These findings suggest that other mutations may be found that contribute significantly to the pathogenesis of hypertension in humans. □

Received 26 September; accepted 29 October 1991.

1. Sutherland, D. J., Ruse, J. L. & Laidlaw, J. C. *Can. med. Assoc. J.* **95**, 1109-1119 (1966).
2. New, M. I. & Peterson, R. E. *J. clin. endocr. Metab.* **27**, 300-305 (1967).
3. Chu, M. D. & Ulick, S. J. *J. biol. Chem.* **257**, 2218-2224 (1982).
4. Ulick, S., Chu, M. D. & Land, M. J. *J. biol. Chem.* **258**, 5498-5502 (1983).
5. Gomez-Sanchez, C. E. *et al. J. clin. endocr. Metab.* **59**, 1022-1024 (1984).
6. Ulick, S. *et al. J. clin. endocr. Metab.* **71**, 1151-1157 (1990).
7. Mornet, E., Dupont, J., Vitek, A. & White, P. C. *J. biol. Chem.* **264**, 20961-20967.
8. Kawamoto, T. *et al. Biochem. biophys. Res. Commun.* **173**, 309-316 (1990).
9. Ogishima, T. *et al. J. biol. Chem.* **266**, 10731-10734 (1991).
10. Chua, S. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 7193-7197 (1987).
11. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. *Genomics* **5**, 874-879 (1989).
12. Gilliland, G., Perrin, S., Blanchard, K. & Bunn, F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2725-2729 (1990).
13. Simpson, E. R. & Waterman, M. R. *A. Rev. Physiol.* **50**, 427-440 (1988).
14. White, P. C., Vitek, A., Dupont, B. & New, M. I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4436-4440 (1988).
15. Platt, R. in *The Epidemiology of Hypertension* (eds Stamler, J., Stamler, R. & Pullman, T. N.) 9-17 (Grune & Stratton, New York, 1967).
16. Pickering, G. W. in *The Epidemiology of Hypertension* (eds Stamler, J., Stamler, R. & Pullman, T. N.) 18-27 (Grune & Stratton, New York, 1967).
17. Ward, R. in *Hypertension: Pathophysiology, Diagnosis and Management* (eds Laragh, J. H. & Brenner, B. M.) 81-100 (Raven, New York, 1990).
18. Ulick, S., Chan, C. K. & Wang, J. C. *J. Steroid Biochem. molec. Biol.* **38**, 59-66 (1991).
19. Bell, G., Karam, J. & Rutter, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5759-5763 (1981).
20. Lifton, R. P., Sardet, C., Pouyssegur, J. & Lalouel, J. M. *Genomics* **7**, 131-135 (1990).
21. Feinberg, A. P. & Vogelstein, B. A. *Analyt. Biochem.* **132**, 6-13 (1983).
22. Schowalter, D. B. & Sommer, S. S. *Analyt. Biochem.* **177**, 90-94 (1989).

ACKNOWLEDGEMENTS. We thank E. Hillas for technical assistance, members of K2061 and their physicians for their contribution to this work, and X. Jeunemaitre, J. Stevens, D. Wilson and G. Williams for discussion and critical review of the manuscript. R.P.L. is a Clinician-Scientist of the American Heart Association; J.M.L. is an Investigator of the Howard Hughes Medical Institute. This work was supported by the NIH (R.G.D., S.U. and J.M.L.).

Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes

B. Palenik & R. Haselkorn

Molecular Genetics and Cell Biology, University of Chicago,
920 E. 58th Street, Chicago, Illinois 60637, USA

PROCHLOROPHYTES are prokaryotes that carry out oxygenic photosynthesis using chlorophylls *a* and *b*, but lack phycobiliproteins as light-harvesting pigments¹. These characteristics distinguish them from cyanobacteria, which contain phycobiliproteins, but no chlorophyll *b*. Three prochlorophyte genera have been described: *Prochloron*¹⁻³, *Prochlorothrix*⁴ and *Prochlorococcus*^{5,6}. The prochlorophytes share their pigment characteristics with green plant and euglenoid chloroplasts, which has led to a debate on whether these chloroplasts may have arisen from an endosymbiotic prochlorophyte rather than a cyanobacterium^{2,7}. Molecular sequence data, including those presented here based on a fragment of the *rpoC1* gene encoding a subunit of DNA-dependent RNA polymerase, indicate that the known prochlorophyte lineages do not include the direct ancestor of chloroplasts⁸⁻¹¹. We also show that the prochlorophytes are a highly diverged polyphyletic group. Thus the use of chlorophyll *b* as a light-harvesting pigment has developed independently several times in evolution. Similar conclusions have been reached in parallel studies using 16S ribosomal RNA sequences¹².

The γ subunit of cyanobacterial RNA polymerases^{13,14} is homologous to the chloroplast RNA polymerase C1 subunit, consistent with the theory that a cyanobacterium-like organism gave rise to the chloroplast¹⁵. On the basis of available *rpoC1*

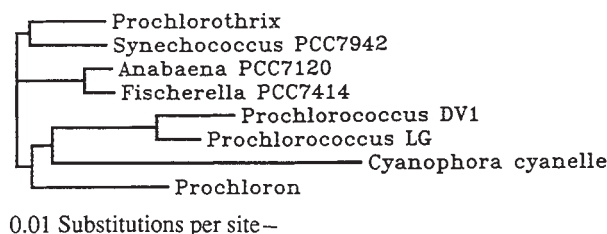


FIG. 2 Unrooted distance matrix method phylogenetic tree of photosynthetic prokaryotes and the *Cyanophora* cyanelle. Distances are proportional to horizontal branch lengths, in units of nucleotide substitutions per site.

METHODS. Using parts of the aligned amino-acid sequences in Fig. 1 (overlined by +++), the corresponding first and second codon positions were used in pairwise comparisons to calculate corrected evolutionary distances (Jukes-Cantor model) using DNADIST of PHYLIP 3.2 (ref. 21). A Fitch-Margoliash distance-matrix method, FITCH of PHYLIP 3.2, with the Global search option, was then used to derive the phylogenetic tree shown. The topology of the tree was not changed by calculating distances using the model of Kimura or by omitting the cyanelle from the analysis.

sequence data, we developed polymerase chain reaction (PCR) primers for obtaining prochlorophyte and cyanobacterial *rpoC1* gene sequences, even in the presence of contaminating DNAs. These sequences can be used for understanding the evolutionary radiation of the cyanobacteria-chloroplast lineage while still allowing the resulting phylogenies to be rooted in a universal tree¹⁶.

We were particularly interested in using *rpoC1* gene sequences to investigate whether prochlorophytes are a monophyletic group within the photosynthetic prokaryotes. In addition, did one or more prochlorophytes branch more closely to the green chloroplasts than the *Cyanophora paradoxa* cyanelle? The cyanelle is a photosynthetic organelle that resembles an endosymbiotic cyanobacterium complete with phycobiliproteins¹⁷. 16S rRNA data suggest that the cyanelle is the closest known relative to the ancestor of the chloroplast¹⁸. If an endosymbiotic prochlorophyte rather than an endosymbiotic cyanobacterium gave rise to the chloroplast, that prochlorophyte lineage should branch more closely to the plant chloroplasts than the cyanelle does.

Analysis of the nucleotide sequence data corresponding to aligned amino-acid positions (overlined by +) in Fig. 1 for the photosynthetic prokaryotes and the cyanelle indicate that the prochlorophytes are highly diverged (Fig. 2). On the basis of calculations of evolutionary distance using the Jukes-Cantor model, for example, the prochlorophytes are as divergent as *Synechococcus* PCC7942 (*Anacystis nidulans*) and *Anabaena* PCC7120, two of the most highly divergent cyanobacteria known, on the basis of 16S rRNA sequence data¹⁸.

Branching patterns of the cyanobacteria and prochlorophytes inferred using distance matrix (Fig. 2) and maximum likelihood methods (not shown) further demonstrate the prochlorophytes are not a monophyletic group. *Prochlorothrix* groups with *Synechococcus* PCC7942, consistent with its position in trees inferred from 16S rRNA (ref. 8), *psbA* (refs 19,20) and RuBP carboxylase/oxygenase⁹, whereas the two *Prochlorococcus* isolates are a sister group with the *C. paradoxa* cyanelle (Fig. 2). A bootstrap DNA parsimony analysis^{21,22} was used to generate a majority rule consensus tree (Fig. 3). This tree has a similar branching pattern to that of Fig. 2, but it suggests that the relative branching of *Prochloron*, *Prochlorothrix* and *Synechococcus* PCC7942 is uncertain. There was no evidence from the replicate trees generated by the bootstrap that trees with the prochlorophytes as a monophyletic group were a viable alternative.

Interestingly, the two *Prochlorococcus* isolates, *Prochlorococcus marinus* (LG) and *Prochlorococcus* sp. (DV1) group together (Figs 2 and 3) in all analyses, but their sequences are more divergent than *Anabaena* PCC7120 (a heterocyst-forming

cyanobacterium) is from *Fischerella* PCC7414 (a branching heterocyst-forming cyanobacterium).

The relationships among the photosynthetic prokaryotes, cyanelles, chloroplasts, and eubacteria were investigated using a more restricted data set (Fig. 4). This distance matrix method tree reproduces the phylogeny of the chloroplasts and cyanelle inferred from 16S rRNA analysis^{8,18}. None of the known pro-

chlorophytes grouped with the chloroplasts to the exclusion of the cyanelle. In this smaller data set, the specific branching patterns of the prochlorophytes is not significantly defined. In a bootstrap DNA parsimony analysis, the cyanelle branched with the chloroplasts in 80% of the replicates, whereas in the other 20% of the replicates it grouped with the cyanobacteria-prochlorophyte line. In no replicate did a prochlorophyte group

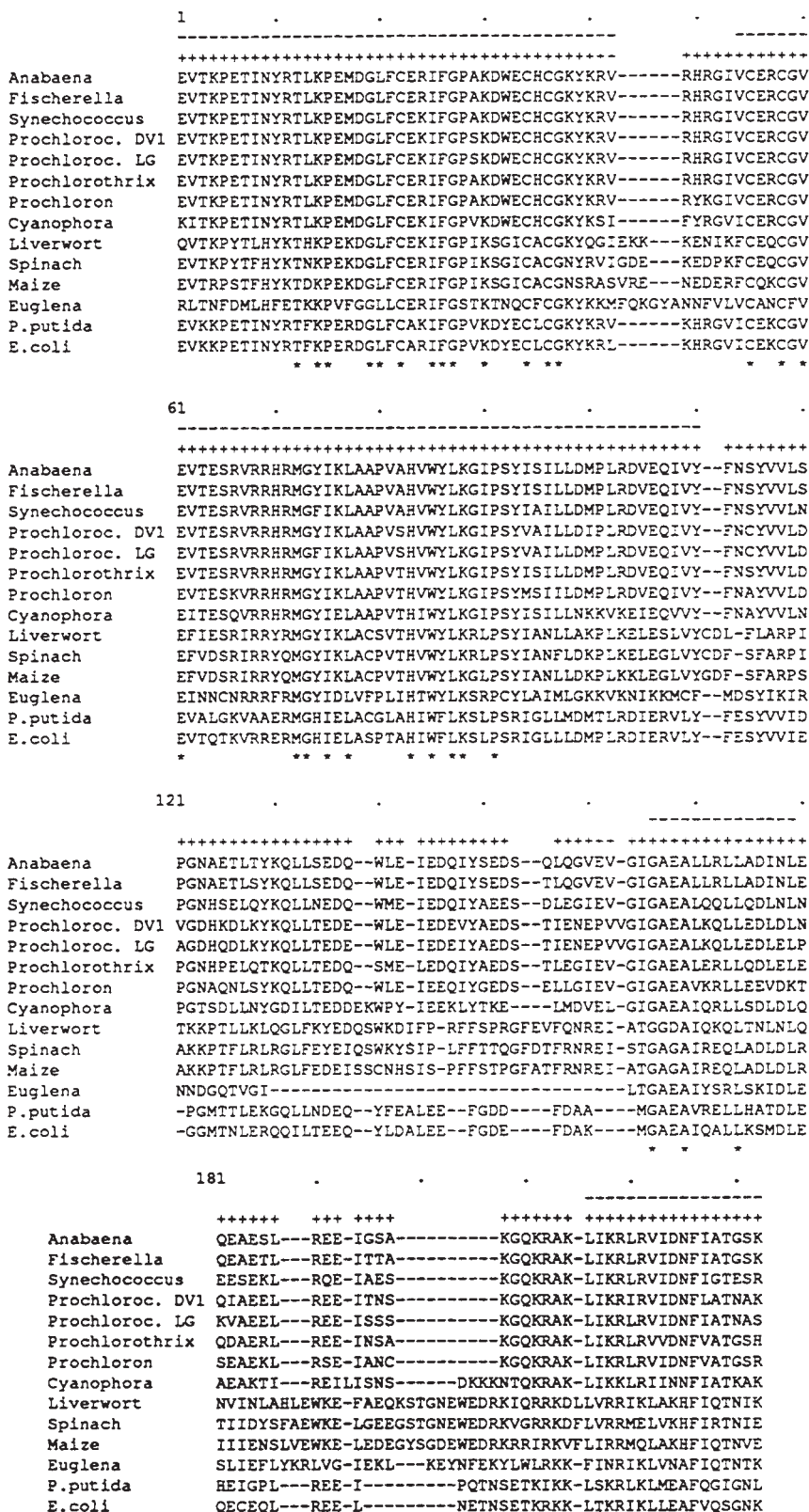


FIG. 1 Aligned predicted amino-acid sequences from a fragment of DNA-dependent RNA polymerases from photosynthetic prokaryotes, a cyanelle, chloroplasts, and two eubacteria (single-letter amino-acid code). ***, Conserved amino acid; +++, data used for Figs 2 and 3; ---, data used for Fig. 4.

METHODS. *Anabaena* PCC7120 (ref. 15); maize²³, spinach²⁴, liverwort²⁵, *Escherichia coli*²⁶, *Pseudomonas putida*²⁷, *Euglena* (Yepiz-Plascencia, G., Radebaugh, C. & Hallick, R. B., unpublished data, GenBank accession number X17191) and predicted sequences from other organisms were aligned using CLUSTAL²⁸ with some corrections by hand. DNA from organisms investigated in this study was obtained from others as noted in the acknowledgements or isolated from cells in culture by standard procedures. Cultures of the *Prochlorococcus* spp. and *Prochlorothrix hollandica* contained nonphotosynthetic bacteria, but are unialgal. For PCR amplification, each of four primers with a different nucleotide at position 22 (5' GGTACCNAAYGNSARRTNGTNGG 3') was paired with a second primer (5' GAGCTCYAWNACCATCC-AYTCNGG 3'). Reactions were run in a Perkin-Elmer Cetus thermal cycler using the reagents and suggested concentrations from an AmpliTaq kit (Perkin Elmer). Typical cycling conditions were 92 °C (1.5 min), 50 °C (1 min.), 72 °C (1 min) for 35 cycles. PCR products were cloned into pUC19 or the TA cloning vector (Invitrogen) using recommended procedures. One cloned PCR product from each organism was sequenced (both strands) using double-stranded dideoxy sequencing (Sequenase, USB). DNA from several organisms was amplified, cloned and sequenced more than once to show that the PCR error rate was less than one error per sequence. The *rpoC1* gene fragment sequence of the *Cyanophora* cyanelle was obtained using standard techniques. The sequences have been deposited with EMBL under the accession numbers Z11152-55 and Z11159-61.

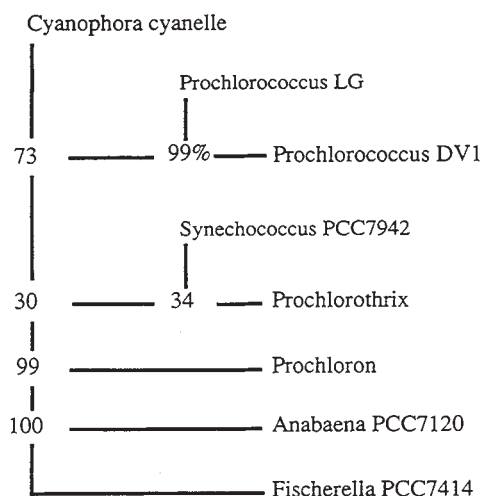


FIG. 3 Unrooted majority-rule consensus tree of photosynthetic prokaryotes and the *Cyanophora* cyanelle derived using a bootstrap DNA parsimony analysis of the DNA sequence data used to derive Fig. 2. Analysis used DNABOOT from PHYLIP 3.2. The numbers at the nodes refer to the percentage of replicates in which the species above and to the right of the node grouped together.

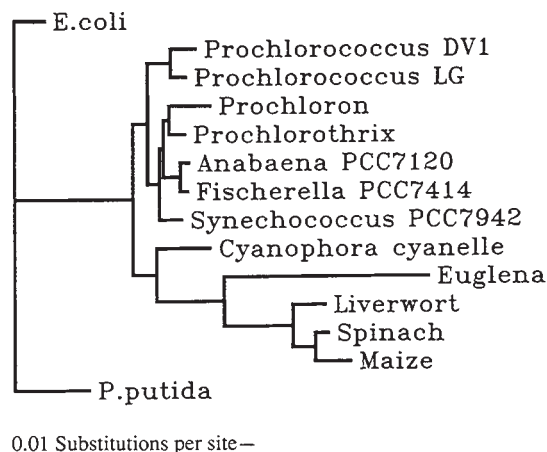


FIG. 4 Unrooted distance matrix method phylogenetic tree of prokaryotes, chloroplasts, and cyanelle derived as for Fig. 2, but using a smaller subset of the data (dotted line in Fig. 1). Distances, in units of nucleotide substitutions per site, are proportional to horizontal branch length.

with the chloroplast lineage to the exclusion of the cyanelle.

We cannot exclude the possibility that an as yet undiscovered prochlorophyte lineage was directly related to the ancestor of the chloroplast, but these results suggest that the plant chloroplast is derived from an independent development (or horizontal gene transfer) of the use of chlorophyll *b* as a light-harvesting pigment. Thus it seems that this prochlorophyte-plant chloroplast phenotype has arisen independently four (or more) times. □

Received 15 August; accepted 14 October 1991.

- Lewin, R. A. *Nature* **261**, 697-698 (1976).
- Lewin, R. A. & Cheng, L. (eds) *Prochloron, a Microbial Enigma* (Chapman & Hall, New York, 1989).
- Lewin, R. A. *Phycologia* **14**, 153-160 (1975).
- Burger-Wiersma, T., Veenhuis, M., Korthals, H. J. Van de Wiel, C. C. M. & Mur, L. R. *Nature* **320**, 262-264 (1986).
- Chisholm, S. W. *et al.* *Nature* **334**, 340-343 (1988).
- Chisholm, S. W. *et al.* *Arch. Mikrobiol.* (in the press).
- Cavaller-Smith, T. *Biol. J. Linn. Soc. Lond.* **17**, 289-306 (1982).
- Turner, S., Burger-Wiersma, T., Giovannoni, S. J., Mur, L. R. & Pace, N. R. *Nature* **337**, 380-382 (1989).
- Morden, C. W. & Golden, S. S. *J. molec. Evol.* **32**, 379-395 (1991).
- Kishino, H., Miyata, T. & Hasegawa, M. *J. molec. Evol.* **31**, 151-160 (1990).

- Stackebrandt, E. in *Prochloron, a Microbial Enigma* (eds Lewin, R. A. & Cheng, L.) 65-69 (Chapman & Hall, New York, 1989).
- Urbach, E., Robertson, D. & Chisholm, S. W. *Nature* **355**, 267-270 (1992).
- Schneider, G. J., Turner, N. E., Richaud, C., Borbely, G. & Haselkorn, R. *J. biol. Chem.* **262**, 14633-14639 (1987).
- Schneider, G. J. & Haselkorn, R. *J. Bact.* **170**, 4136-4140 (1988).
- Bergsland, K. J. & Haselkorn, R. *J. Bact.* **173**, 3446-3455 (1991).
- Puehler, G. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 4569-4573 (1989).
- Jaynes, J. M. & Vernon, L. P. *Trends Biochem. Sci.* **7**, 22-24 (1982).
- Giovannoni, S. J. *et al.* *J. Bact.* **170**, 3584-3592 (1988).
- Morden, C. W. & Golden, S. S. *Nature* **337**, 382-384 (1989).
- Morden, C. W. & Golden, S. S. *Nature* **339**, 400 (1989).
- Felsenstein, J. *Cladistics* **5**, 164-166 (1989).
- Felsenstein, J. *Evolution* **39**, 783-791 (1985).
- Igloi, G. L., Meinke, A., Döry, I. & Kössel, H. *Molec. Gen. Genet.* **221**, 379-394 (1990).
- Hudson, G. S., Holton, T. A., Whitfield, P. R. & Bottomley, W. J. *J. molec. Biol.* **200**, 639-654 (1988).
- Ohyama, K. *et al.* *Nature* **322**, 572-574 (1986).
- Ovchinnikov, Y. A. *et al.* *Nucleic Acids Res.* **10**, 4035-4044 (1982).
- Borodin, A. M. *et al.* *Bioorg. Khim.* **14**, 1179-1182 (1988).
- Higgins, D. G. & Sharp, P. M. *Gene* **73**, 237-244 (1989).

ACKNOWLEDGEMENTS. We thank B. Brahamsha, J. B. Waterbury and H. Swift for comments on the manuscript, G. Olsen for comments on the manuscript, data alignment and phylogenetic analyses, W. Buikema for use of his program DNALYSIS, and K. Bergsland for prepublication use of *Anabaena* 7120 rpoC1 sequence data. Partially purified *Prochloron* DNA isolated from cells expressed from the ascidian *Lissoclinum patella* was provided by H. Swift. An *EcoRI* fragment containing the rpoC1 subunit of the *Cyanophora paradoxa* cyanelle and cloned into pUC19 by V. L. Stirewalt was provided by D. Bryant. Lyophilized cells of *Fischerella* PCC7414 were provided by J. B. Waterbury. Cells from *Prochlorococcus* sp. isolate DV1 (isolated by D. Vaulot) were provided by S. Frankel. This work was supported by a NSF Marine Biotechnology postdoctoral fellowship (to B.P.).

Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation

Ena Urbach, Deborah L. Robertson* & Sallie W. Chisholm

Ralph M. Parsons Laboratory,
48-425 Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

* Department of Molecular Genetics and Cellular Biology,
University of Chicago, Chicago, Illinois 60637, USA

THE taxonomic group Prochlorales (Lewin 1977) Burger-Wiersma, Stal and Mur 1989 was established to accommodate a set of prokaryotic oxygenic phototrophs which, like plant, green algal and euglenoid chloroplasts, contain chlorophyll *b* instead of phycobiliproteins. Prochlorophytes were originally proposed (with concomitant scepticism¹⁻³) to be a monophyletic group sharing a common ancestry with these 'green' chloroplasts⁴⁻⁶. Results from molecular sequence phylogenies, however, have suggested that *Prochlorothrix hollandica*⁷ is not on a lineage that leads to plastids⁸⁻¹². Our results from 16S ribosomal RNA sequence comparisons, which include new sequences from the marine picoplankter *Prochlorococcus marinus*^{13,14} and the *Lissoclinum patella* symbiont *Prochloron* sp.¹⁵, indicate that prochlorophytes are polyphyletic within the cyanobacterial radiation, and suggest that none of the known species is specifically related to chloroplasts. This implies that the three prochlorophytes and the green chloroplast ancestor acquired chlorophyll *b* and its associated structural proteins in convergent evolutionary events. We report further that the 16S rRNA gene sequence from *Prochlorococcus* is very similar to those of open ocean *Synechococcus* strains (marine cluster A (ref. 16)), and to a family of 16S rRNA genes shotgun-cloned from plankton in the north Atlantic and Pacific Oceans¹⁷⁻¹⁹.

The order Prochlorales subsumes two formally described genera: *Prochloron*, a symbiont of marine ascidians²⁰, and *Prochlorothrix*, a free-living, filamentous, fresh-water plankter⁷. Like green chloroplasts, these prokaryotes have appressed thylakoid membranes containing chlorophylls *a* and *b* and lacking phycobilisomes. The recently discovered *Prochlorococcus*, a free-living, tiny unicell that is widespread and abundant in the oceans¹³, shares these traits with its 'relatives', but differs in that it contains divinyl chlorophylls *a* and *b* (exclusively), and α -rather than β -carotene^{14,21}.

TABLE 1 Evolutionary distance and fractional similarity matrix for 16S rRNA sequences from *A. tumefaciens* and members of the prokaryotic oxygenic phototroph radiation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. <i>Agrobacterium tumefaciens</i>		813	798	805	810	808	806	811	802	812	794	812	798	802	797	788	805	808
2. SAR100	215		918	978	887	878	892	889	892	891	895	892	889	892	885	844	985	993
3. <i>Synechococcus</i> PCC6301	236	87		913	883	873	893	881	887	883	895	889	893	911	886	856	919	911
4. <i>Prochlorococcus marinus</i>	226	23	93		886	874	882	887	897	892	892	897	888	898	886	852	975	975
5. <i>Prochloron</i> sp.	219	122	128	123		920	906	885	921	941	903	920	903	885	910	879	883	885
6. <i>Synechocystis</i> PCC6308	221	134	139	138	85		900	854	892	908	885	896	889	866	888	848	868	875
7. <i>Synechocystis</i> PCC6906	225	117	115	129	101	108		863	923	914	899	892	895	883	890	845	885	890
8. <i>Gloeobacter</i> PCC7421	218	120	129	122	125	163	151		883	878	898	905	883	889	884	855	881	885
9. <i>Gloeotheca</i> PCC6501	229	117	122	110	83	117	81	127		914	903	907	895	890	896	862	891	886
10. <i>Myxosarcina</i> PCC7312	216	118	127	117	62	98	91	134	91		906	913	896	892	893	866	886	889
11. <i>Lyngbya</i> PCC7419	242	113	113	117	104	125	109	110	104	101		923	908	892	910	868	892	896
12. <i>Oscillatoria</i> PCC6304	217	117	120	110	85	111	117	102	99	93	81		901	896	897	867	888	893
13. <i>Anabaena</i> PCC7122	236	120	115	121	104	120	113	127	113	112	98	106		876	905	872	891	886
14. <i>Prochlorothrix hollandica</i>	231	116	95	110	125	147	127	121	119	117	117	112	136		883	863	895	889
15. <i>Cyanophora</i> cyanelle	237	125	123	124	96	121	119	126	112	115	96	110	102	128		883	878	881
16. <i>Marchantia</i> chloroplast	249	175	160	165	131	169	173	161	152	147	146	146	141	151	127		846	840
17. <i>Synechococcus</i> WH7805	226	15	86	25	127	145	125	130	118	124	117	121	118	113	133	172		986
18. <i>Synechococcus</i> WH8103	222	7	95	25	125	136	119	124	123	120	112	116	123	120	130	180	14	

Estimated evolutionary distances²⁹ ($\times 1,000$, below the diagonal) and fractional similarity values ($\times 1,000$, above) for pairs of 16S rRNA sequences for *Prochlorothrix hollandica*⁸, other prochlorophytes (this work), cyanobacteria^{22,30} (D. L. Distel and J. B. Waterbury, manuscript in preparation), photosynthetic organelles^{22,31}, a full-length, shotgun-cloned sequence from the Sargasso Sea¹⁹, and *A. tumefaciens*³², used in the construction of the phylogenetic tree in Fig. 1a. Sequences were aligned according to their conserved primary and secondary structures and compared by the computer program of Olsen²³, using 788 nucleotides of sequence unambiguously aligned for all organisms considered. Gap values were set at 0.5 nucleotide substitution. *Marchantia*, *Marchantia polymorpha*; *Cyanophora*, *Cyanophora paradoxa*. Genomic DNA from *Prochlorococcus* (isolated from the Sargasso Sea, 30 May 1988, 28° 58.9' N, 64° 21.5' W, at a depth of 120 m; ref. 14) was extracted using standard techniques³³ from a dilution culture inoculated with an average cell density of one cell per culture (isolate SSW5, 58% probability of clonality). *Prochloron* cells were collected in February 1990 from reef flats in the Kamori Channel, Koror, Palau, West Caroline Islands (7° 25' N, 134° 30' E). Collection and method of isolating the symbiont from the host, *Lissoclinum patella*, are described in ref. 34. DNA was extracted from frozen cells using standard protocols³⁵, except that the extraction buffer was adjusted to 0.25 M EDTA. 16S rRNA genes were amplified from genomic DNA of *Prochlorococcus* using the polymerase chain reaction (PCR) with the primers PLG1.1 (ACGGGTGAGTAACGCGTRA) and PLG2.1 (CTTATGCAGCGAGTTGCAGC), designed to be specific for members of the oxygenic phototroph lineage and thus select against sequences from heterotrophic contaminants in the culture. *Prochloron* sequences were amplified using the primers P1 (AGAGTTTGATCCTGGCTCAG) and P2 (CTTGTACGACTTCACCCC), which amplify 16S rRNA sequences from all eubacteria, as there was no significant contamination from heterotrophs in this preparation. PCR reactions contained 10 ng genomic DNA, 100 nM each primer, 200 μ M each dNTP, 2.25 mM MgCl₂, 50 mM KCl, 10 mM Tris (pH 8.4), and 5 units *Taq* polymerase (Perkin Elmer Cetus). Reaction volumes were 200 μ l, covered with 100 μ l sterile mineral oil. Cycle parameters were: 94 °C for 1 min, *T_d* for 1 min, and 72 °C for 2 min. Samples were processed for 40 to 50 cycles. *T_d* was 66 °C for the PLG1.1/PLG2.1 primer pair, and 58 °C for the P1/P2 pair. PCR products were purified from preparative polyacrylamide gels by electro-elution (D-gel, EpiGene), and reamplified asymmetrically using 2 nM of one limiting primer. Asymmetric PCR products from both coding and complementary strands were ethanol-precipitated twice with ammonium acetate and used for dideoxy-nucleotide sequencing with Sequenase version 2 (US Biochemical Corporation) in parallel reactions containing dGTP and dITP, plus and minus single-strand binding protein. 1,147 bases of continuous sequence (*Escherichia coli* 16S rRNA nucleotide positions 128 to 1,312) were determined for *Prochlorococcus* and 1,377 bases (*E. coli* positions 28 to 1,452) for *Prochloron*. Sequences are available through GenBank, accession numbers X63140 and X63141.

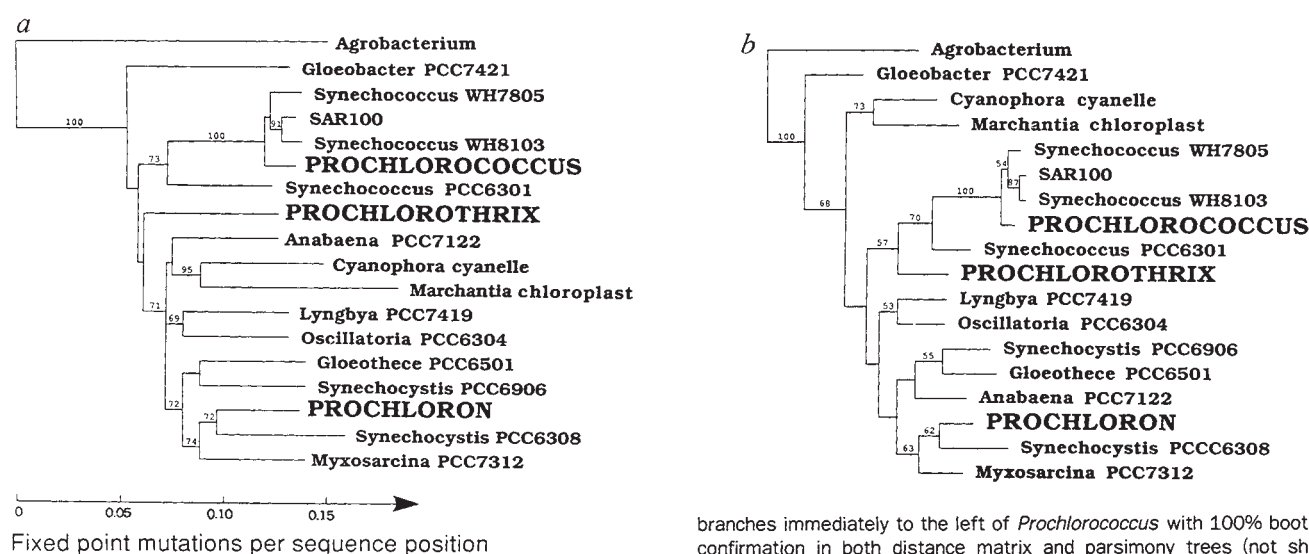


FIG. 1 Nucleic-acid sequence phylogenies, inferred from 16S rRNA data, illustrating the apparent independent appearance of chlorophyll *b*-containing organisms in the cyanobacterial lineage. The phylogenetic position of a full-length, shotgun-cloned 16S rRNA gene from Sargasso Sea plankton¹⁹ (SAR100) is also shown. A sequence for the freshwater *Synechococcus* strain PCC6307 (D. L. Distel and J. B. Waterbury, manuscript in preparation)

branches immediately to the left of *Prochlorococcus* with 100% bootstrap confirmation in both distance matrix and parsimony trees (not shown). Numbers indicate per cent confirmation by bootstrap analysis, summarizing the tree output from 100 resampled datasets, of the phylogenetic group to the right. Only values greater than 50% are labelled. a, Distance matrix analysis. Horizontal distances are proportional to evolutionary distances (Table 1). b, Parsimony analysis of the same data using the heuristic tree search option of PAUP²⁴. An equally parsimonious tree reverses the relative positions of *Prochlorococcus* and *Synechococcus* WH7805.

TABLE 2 Evolutionary distance and fractional similarity matrix for 16S rRNA sequences from *A. tumefaciens*, members of the prokaryotic oxygenic phototroph radiation and shotgun clones from marine plankton communities

	1	2	3	4	5	6	7	8	9	10
1. <i>Agrobacterium tumefaciens</i>		782	808	821	797	812	805	799	812	805
2. SAR6	258		967	954	918	964	912	951	951	971
3. SAR7	222	33		987	925	977	919	964	984	971
4. SAR100	205	47	13		931	977	932	964	984	971
5. <i>Synechococcus</i> PCC6301	236	87	79	72		935	974	922	915	935
6. <i>Prochlorococcus marinus</i>	217	37	23	23	68		942	987	981	994
7. <i>Prochlorothrix hollandica</i>	226	94	86	72	27	61		929	922	942
8. <i>Synechococcus</i> WH7805	234	51	37	37	83	13	75		981	981
9. <i>Synechococcus</i> WH8103	217	51	16	16	90	20	82	20		974
10. ALO37	226	30	30	30	68	7	61	20	26	

Estimated evolutionary distances ($\times 1,000$, below the diagonal) and fractional similarity values ($\times 1,000$, above) for pairs of 16S rRNA sequences for *Prochlorothrix*⁸, other prochlorophytes (this work), cyanobacteria³⁰ (D. L. Distel and J. B. Waterbury, manuscript in preparation), shotgun-cloned 16S rRNA genes¹⁷⁻¹⁹, and *Agrobacterium*³², used in the construction of the phylogenetic tree in Fig. 2. The 186 bases of aligned sequence corresponding to position 306 to 514 in the *E. coli* 16S rRNA sequence were analysed as described in the legend to Table 1.

We compared 16S rRNA gene sequences for both *Prochlorococcus* and *Prochloron* with sequences from the oxygenic phototroph database^{8,22} (D. L. Distel and J. B. Waterbury, manuscript in preparation) using both distance matrix²³ (Fig. 1a; Table 1) and parsimony²⁴ (Fig. 1b) algorithms. Bootstrap resampling²⁵ was used to estimate reliability of the inferred trees (Fig. 1a, b). For simplicity we refer only to the results of the distance matrix analysis in our discussion, although results from parsimony analysis are in essential agreement.

The 16S rRNA tree strongly supports the derivation of both *Prochlorococcus* and green chloroplasts (represented in our analysis by the *Marchantia polymorpha* chloroplast) from different, phycobilisome-containing ancestors among the cyanobacteria (Fig. 1a). *Prochlorococcus* forms a shallowly branching cluster with marine *A. Synechococcus* strains WH7805 and WH8103 (ref. 16) (confirmed in 100% of bootstrap resamplings), before which branch the cyanobacteria *Synechococcus* PCC6301 (D. L. Distel and J. B. Waterbury, manuscript in preparation) (not shown; confirmed in 100% of bootstrap resamplings) and *Synechococcus* PCC6301 (confirmed in 73% of bootstrap resamplings). The *Marchantia* chloroplast branches

with the *Cyanophora paradoxa* cyanelle (95% bootstrap confirmation), in accord with evidence that green chloroplasts and the cyanobacterium-like cyanelle descend from a common ancestor^{22,26}. Branching patterns for *Prochloron* and *Prochlorothrix* are less than 95% confirmed, thus their descent from separate, phycobilisome-containing ancestors, though likely, is less certain than for *Prochlorococcus* and the chloroplasts.

Prochlorococcus is closely related not only to the open-ocean strains, *Synechococcus* WH7805 and WH8103, but also to a diversity of shotgun-cloned 16S rRNA gene sequences from the Sargasso Sea^{17,19} and the north Pacific Ocean¹⁸ (Figs 1 and 2; Tables 1 and 2). Individual samplings from these sites, where both *Prochlorococcus* and *Synechococcus* are common, have routinely netted a puzzling multiplicity of sequences, exhibiting greater than 95% similarity, interpreted as belonging to open-ocean *Synechococcus*¹⁷⁻¹⁹. This diversity can now be partially attributed to sympatric populations of *Synechococcus* and *Prochlorococcus*, which, despite their sequence similarity, are separate species by ecological as well as phenotypic criteria²⁷.

We conclude that the phycobilisome⁻/chlorophyll *b*⁺ (green chloroplast) phenotype seems to have evolved *de novo* at least three, and possibly four times during the evolution of cyanobacteria: once in the ancestry of the green chloroplast lineage, again in that of *Prochlorococcus*, and once or twice in the ancestries of *Prochloron* and *Prochlorothrix*, the mutual independence of which has not positively been demonstrated. The shallowness of the *Prochlorococcus* cluster suggests a particularly recent origin for this organism's pigment phenotype. In the light of these results, and those of Palenik and Haselkorn²⁸, the order Prochlorales can no longer be justified as a natural grouping. The taxon should therefore be abandoned, and prochlorophytes reclassified in the cyanobacteria. □

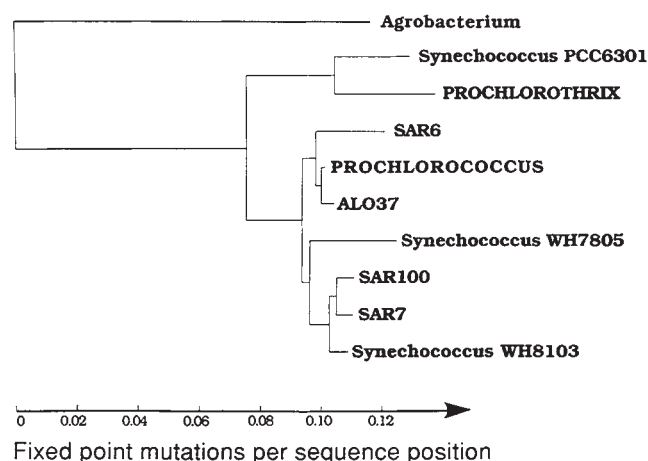


FIG. 2 Phylogenetic tree illustrating relationships among prochlorophytes, members of the *Synechococcus* group and shotgun-cloned sequences from Sargasso Sea plankton^{17,19} (SAR) and Pacific Ocean picoplankton¹⁸ (ALO). The tree was constructed using distance matrix analysis and 16S rRNA sequence data (Table 2). Differences in the branching order between this tree and the one shown in Fig. 1a are presumably due to the short length of the shotgun-cloned sequences, which limits the resolution of this analysis. Bootstrap analysis indicates that *Prochlorococcus*, *Synechococcus* strains WH8101, WH7805, and the cloned sequences form a coherent phylogenetic cluster distinct from *Prochlorothrix* and *Synechococcus* PCC6301, but the branching order in this cluster is indeterminate.

Received 27 August; accepted 18 October 1991.

1. Cavalier-Smith, T. *Biol. J. Linn. Soc. Lond.* **17**, 289-306 (1982).
2. Whitton, B. A. & Carr, N. G. in *The Biology of the Cyanobacteria* (eds Carr, N. G. & Whitton, B. A.) 1-8 (University of California, Berkeley, 1982).
3. Lewin, R. A. & Cheng, L. in *Prochloron: A Microbial Enigma* (eds Lewin, R. A. & Cheng, L.) 1-7 (Chapman and Hall, New York, 1989).
4. Van Valen, L. M. & Maierana, V. C. *Nature* **287**, 248-250 (1980).
5. Lewin, R. A. *Ann. N. Y. Acad. Sci.* **361**, 325-329 (1981).
6. Margulis, L. *Symbiosis in Cell Evolution* (Freeman, San Francisco, 1981).
7. Burger-Wiersma, T. et al. *Int. J. Syst. Bacteriol.* **39**, 250-257 (1989).
8. Turner, S. et al. *Nature* **337**, 380-382 (1989).
9. Morden, C. W. & Golden, S. S. *Nature* **337**, 382-385 (1989).
10. Morden, C. W. & Golden, S. S. *Nature* **339**, 400 (1989).
11. Morden, C. W. & Golden, S. S. *J. molec. Evol.* **32**, 379-395 (1991).
12. Kishino, H., Miyata, T. & Hasegawa, M. *J. molec. Evol.* **31**, 151-160 (1990).
13. Chisholm, S. W. et al. *Nature* **334**, 340-343 (1988).
14. Chisholm, S. W. et al. *Archiv. Microbiol.* (in the press).
15. Lewin, R. A. & Withers, N. W. *Nature* **256**, 735-737 (1975).
16. Waterbury, J. B. & Rippka, R. in *Bergey's Manual of Systematic Bacteriology* Vol. 3 (eds Stanley, J. T. et al.) 1728-1746 (Williams and Wilkins, Baltimore, 1989).
17. Giovannoni, S. J. et al. *Nature* **345**, 60-62 (1990).
18. Schmidt, T. M., DeLong, E. F. & Pace, N. R. *J. Bact.* **173**, 4371-4378 (1991).
19. Britschgi, T. B. & Giovannoni, S. J. *Appl. envir. Microbiol.* **57**, 1707-1713 (1991).
20. Lewin, R. A. *Phycologia* **1**, 217 (1977).

21. Goericke, R. & Repeta, D. *Limnol. Oceanogr.* (in the press).
22. Giovannoni, S. J. *et al. J. Bact.* **170**, 3584–3592 (1988).
23. Olsen, G. J. *Meth. Enzym.* **164**, 793–838 (1988).
24. Swofford, D. L. *Paup* Version 3.0 (Illinois Natural History Survey, Champaign, Illinois, 1989).
25. Feisenstein, J. *Evolution* **39**, 783–791 (1985).
26. Douglas, S. E. & Turner, S. J. *molec. Evol.* **33**, 267–273 (1991).
27. Olson, R. J., Chisholm, S. W., Zettler, E. R., Altabet, M. A. & Dusenberry, J. A. *Deep Sea Res.* **37**, 1033–1051 (1990).
28. Palenik, B. & Haselkorn, R. *Nature* **355**, 265–267 (1992).
29. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21 (Academic, New York, 1969).
30. Tomioka, N. & Sugiyama, M. *Molec. gen. Genet.* **191**, 45–50 (1983).
31. Ohyama, K. *et al. Pl. molec. Biol. Reporter* **4**, 148–175 (1986).
32. Yang, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 4443–4447 (1985).
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
34. Swift, H. & Robertson, D. L. *Symbiosis* **10**, 95–113 (1991).
35. Dzelzkalns, V. A., Szekeres, M. & Mulligan, B. J. in *Plant Molecular Biology* (ed. Shaw, C. H.) 277–299 (IRL, Washington DC, 1988).

ACKNOWLEDGEMENTS. We thank M. L. Sogin for assistance with phylogenetic analyses, D. L. Distel and J. B. Waterbury for unpublished sequences, and W. Thilly and P. Keohavong for technical support. This work was supported by grants from the NSF, EPA and Office of Naval Research. D.L.R. was supported by a training grant from the NIH.

Structural basis of latency in plasminogen activator inhibitor-1

James Mottonen*, Arne Strand*, Jindrich Symersky*, Robert M. Sweet†, Dennis E. Danley‡, Kieran F. Geoghegan‡, Robert D. Gerard*§ & Elizabeth J. Goldsmith*||

Departments of * Biochemistry and § Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9050, USA

† Department of Biology, Brookhaven National Laboratory, Upton, New York 11973, USA

‡ Department of Molecular Genetics and Protein Chemistry, Central Research Division, Pfizer Inc., Eastern Point Road, Groton, Connecticut, 06340, USA

HUMAN plasminogen activator inhibitor-1 (PAI-1)^{1,2} is the fast-acting inhibitor of tissue plasminogen activator and urokinase³ and is a member of the serpin family of protease inhibitors⁴. Serpins normally form complexes with their target proteases that dissociate very slowly as cleaved species and then fold into a highly stable inactive state⁵ in which the residues that flank the scissile bond (P1 and P1') are separated by about 70 Å (refs 7–9). PAI-1 also spontaneously folds into a stable¹⁰ inactive state without cleavage; this state is termed 'latent' because inhibitory activity can be restored through denaturation and renaturation^{2,10}. Here we report the structure of intact latent PAI-1 determined by single-crystal X-ray diffraction to 2.6 Å resolution. The three-dimensional structure reveals that residues on the N-terminal side of the primary recognition site are inserted as a central strand of the largest β sheet, in positions similar to the corresponding residues in the cleaved form of the serpin α₁-proteinase inhibitor (α₁-PI)⁷. Residues C-terminal to the recognition site occupy positions on the surface of the molecule distinct from those of the corresponding residues in cleaved serpins^{7–9} or in the intact inactive serpin homologue, ovalbumin¹¹, and its cleavage product, plakalbumin¹². The structure of latent PAI-1 is similar to one formed after cleavage in other serpins, and the stability of both latent PAI-1 and cleaved serpins may be derived from the same structural features^{13,14}.

The structure of bacterially expressed latent PAI-1 was solved by molecular replacement using the structure of α₁-PI as a search model, and refined with models based on α₁-PI, plakalbumin and ovalbumin (Fig. 1). The R-factor of the present model is 19.8% on data between 6.0 and 2.6 Å resolution (Table 1). The electron density map clearly shows the conformation of the reactive centre loop and the quality of the density in sheet

TABLE 1 Crystallographic data and refinement

Crystal data		Refinement parameters	
Space group	C2	Protein atoms	2946
Unit cell (Å)		Δ bond (Å)	0.014
a	157.59	Δ angle (degrees)	3.8
b	47.84	Resolution (Å)	6.0–2.6
c	62.88	Reflections	9,332
β	107.78	R _{cryst} [†]	0.198
Resolution (Å)	2.6	Superpositions	
Total measurements	44,142	PAI-1–α ₁ -PI (r.m.s. Å)	2.48
Unique reflections	12,239	(348 Cα atoms‡)	
% Complete 25–2.6 Å	87.4	PAI-1–α ₁ -PI	1.68
% Complete 2.7–2.6 Å	64.0	(321 Cα atoms§)	
R _{sym} [*]	0.08	PAI-1–α ₁ -PI	0.91
Rotation function		(114 Cα atoms,	
Rotation peak (r.m.s.)	6.3	β-sheets A, B and C)	
Translation (r.m.s.)	5.7		

* $R_{sym} = \{ \sum_{hkl} \sum_n ((I_{hkl}^n - I_{hkl}) / I_{hkl}) \} / N_{hkl}$ where I_{hkl}^n is the n th measurement of I_{hkl} .

† $R_{cryst} = \sum_{hkl} [|F(obs)_{hkl}| - |F(calc)_{hkl}|] / |F(obs)_{hkl}|$, where $|F(obs)|$ and $|F(calc)|$ are the observed and calculated structure factor amplitudes.

‡ Calculated in 'O' {23} (ref. 25).

§ Found with LSQ-Improve in 'O' and omits residues 81–95 (D-helix), 107–108, 128, 141, 207, 215–216, 278–279, 390–391.

A (Fig. 1). Latent PAI-1 resembles other serpins in overall fold. The core β sheets A, B, and C (Fig. 2a) superimpose fairly closely with those of α₁-PI, whereas the helical segments diverge (Fig. 2a; Table 1). But the placement and conformation of the reactive centre loop, residues Ser 343–Arg 368 (α₁-PI numbering⁴; residues P16–P10'), is different. Thirteen residues, from Ser 343–Val 355 (P16–P4), are incorporated as a central β strand (strand 4) in sheet A (Fig. 2a), as are fourteen residues (P16–P3) in cleaved α₁-PI. This β strand forms many hydrogen bonds and hydrophobic interactions with strands 3A and 5A, which are likely to contribute to the stability of the latent structure.

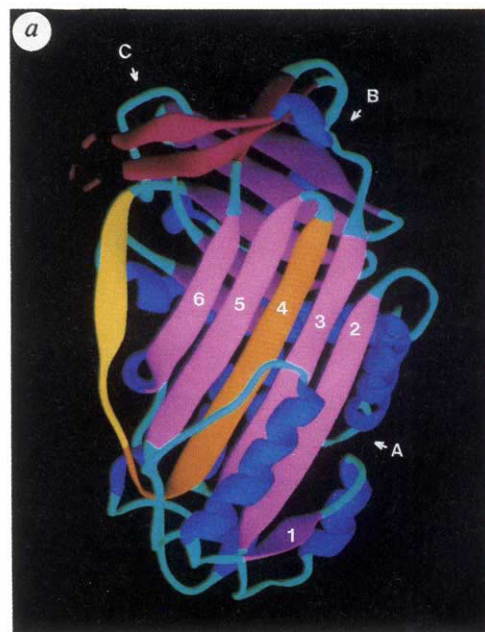
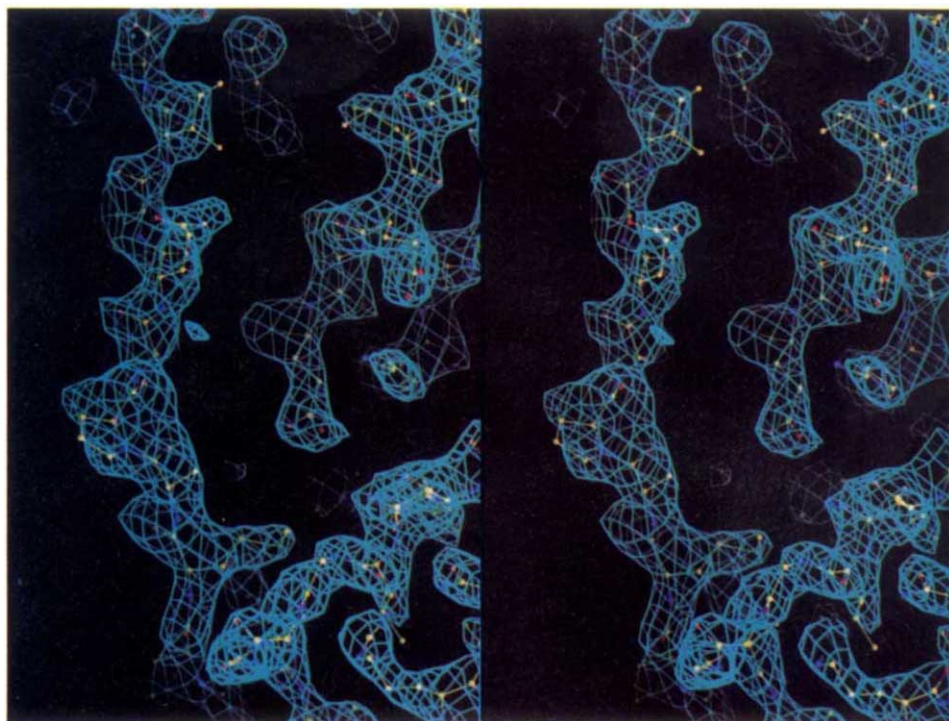


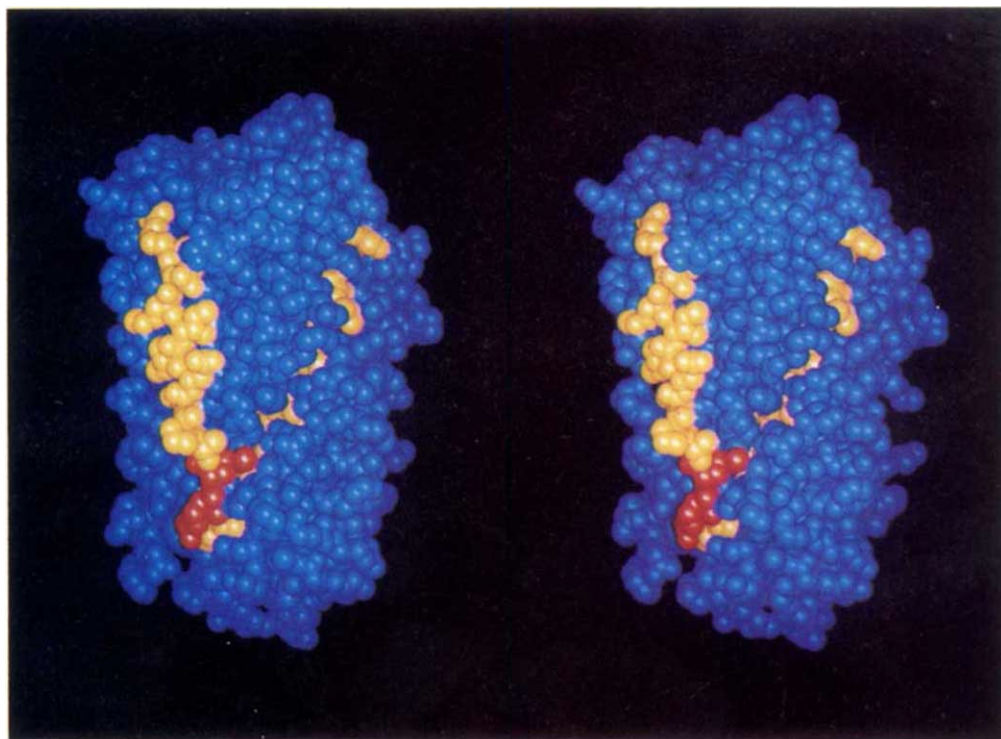
FIG. 1 Stereodiagram of electron density map contoured at 1.2σ of reactive centre loop (Met 358–Ile 364, P1–P6', on the left) and part of sheet A calculated with model-derived phases and coefficients $2|F(\text{obs})| - |F(\text{calc})|$ for data from 6.0–2.6 Å and coefficients $|F(\text{obs})|$ for data from 20.0–6.0 Å (drawn by 'O' (ref. 25) on an Evans and Sutherland ESV workstation). The protein was expressed in *Escherichia coli*, then purified²⁶ and crystallized²⁷ as described. The space group is $C2$, which is higher symmetry than originally reported²⁷. N-terminal sequence analysis of dissolved crystals gave a single polypeptide chain. Native data were collected to 2.8 Å with a Xuong–Hamlin area detector²⁸ and Rigaku rotating anode source and to 2.6 Å resolution with a FAST area detector at the National Synchrotron Light Source at Brookhaven National Laboratories (Table 1). The structure was solved by molecular replacement with α_1 -PI (Protein Data Bank field 6API) as the search model. Solutions to the Crowther rotation²⁹ and translation functions³⁰ were found with MERLOT³¹. The correct translation peak was located by fine-sampling of the rotation function around the observed solution to the rotation function. Models were constructed based on the coordinates of α_1 -PI, plakalbumin (coordinates by courtesy of T. Wright) and ovalbumin (coordinates by courtesy of A. Leslie). Refinement of the models was carried out in XPLOR³² using simulated annealing. The model based on ovalbumin and plakalbumin could not be refined below an R -factor of 25%, and at that stage maps calculated with coefficients $(2|F(\text{obs})| - |F(\text{calc})|)$ revealed strand-to-strand and helix-to-



strand connections corresponding to those of α_1 -PI. A revised model based on α_1 -PI connectivity refined further, and cycles of model building and refinement allowed the remainder of the chain to be traced, except residues Phe 208 to Ser 214. The current model includes 50 water molecules and individual B -factors have been refined with restraints on the B -factors of bonded atoms. 340 out of 348 defined non-glycine residues have (ϕ, ψ) angles within allowed hard-sphere limits.

b

FIG. 2a, Diagram of secondary structure of PAI-1 drawn by RIBBON³³. The three β sheets are coloured pink (A), purple (B) and burgundy (C). The reactive centre loop is coloured orange between Ser 343–Arg 358 (P16–P1), and yellow between Met 359–Arg 368 (P1'–P10'). Disordered residues Phe 208–Ser 214 (part of 3C–4C in α_1 -PI) are dashed. *b*, Solid rendering of latent PAI-1 drawn by 'O'. Reactive centre loop residues are coloured orange, except for Arg 358 and Met 359 (P1, P1'), which are red.



a

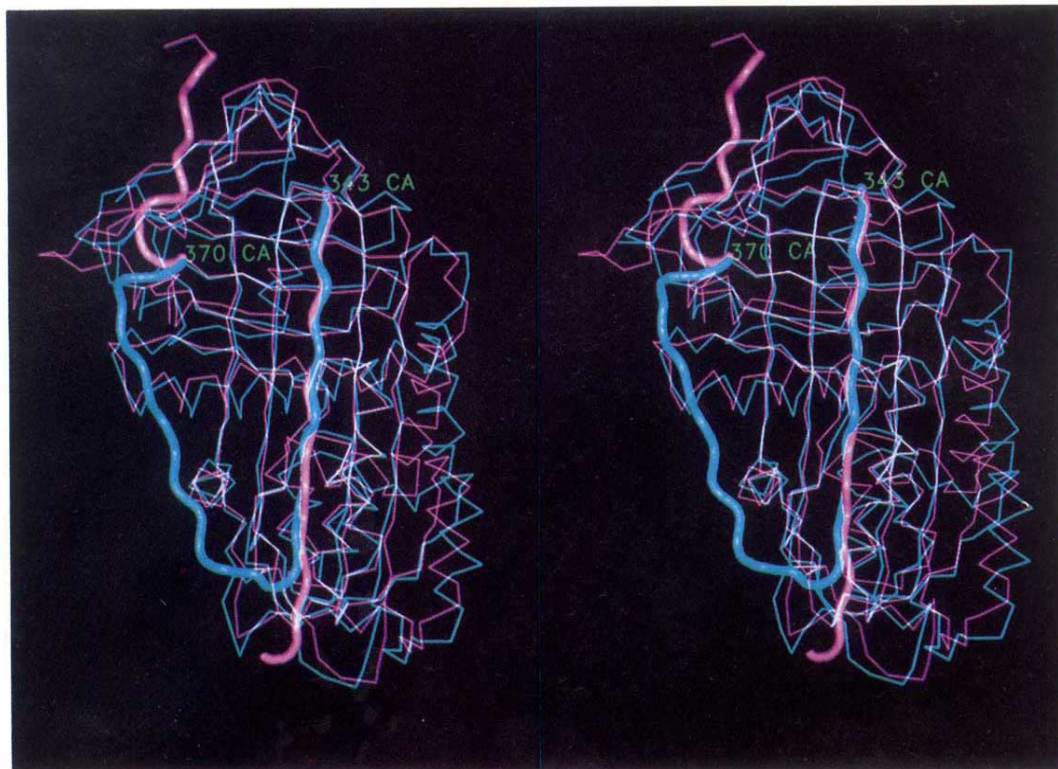
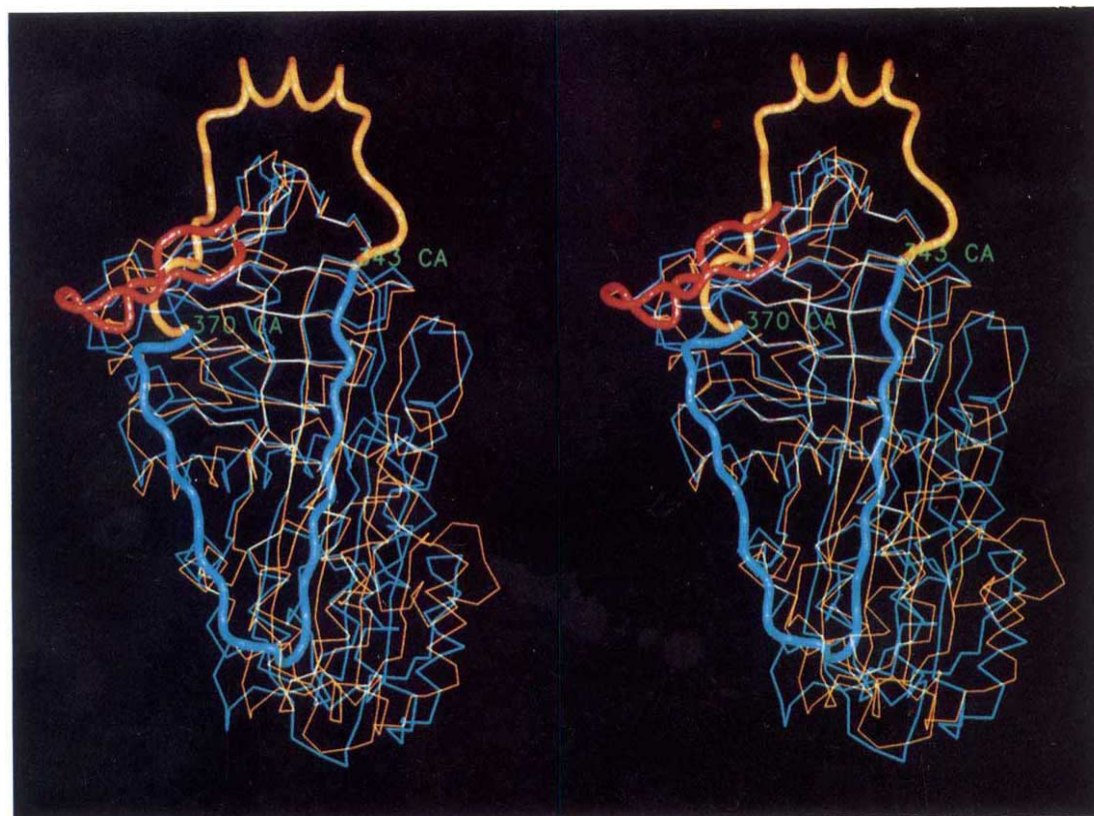


FIG. 3a, Stereosuperposition of latent PAI-1 (cyan) and α_1 -PI (magenta) drawn by 'O'; reactive centre loops 343-368 (P16-P10') are solid. b, Stereosuperposition of latent PAI-1 (cyan) with ovalbumin (gold) drawn by 'O'; reactive centre loops are solid, and residues 202-220 (strands 3C-4C) of ovalbumin are solid red.



The remainder of the reactive centre loop Ser 356-Arg 368 (P2-P10') emerges from sheet A and forms an unusual extended loop which is stretched along the surface of the protein between the C terminus of strand 4A and the N terminus of strand 4B (Fig. 2b). By contrast, most of the corresponding residues in α_1 -PI form the first strand of a β sheet at the top of the molecule (sheet C in Fig. 3a). No backbone hydrogen bonds are evident in the extended loop of latent PAI-1; the residues have raised

temperature factors, but all are visible in the electron density map. Arg 358 and Met 359 (P1 and P1') at the plasminogen activator primary recognition site, and other residues in the extended loop, project on alternate sides of the polypeptide backbone towards the surface of the molecule (Fig. 2b). With the exception of Glu 362 (P4' site), which forms an ion pair with Arg 48, residues in the extended loop make few contacts with the rest of the protein. This suggests that these residues do

not contribute significantly to the stability of the latent structure. Apparently the extended loop is driven into this taut conformation by the formation of strand 4A.

Latent PAI-1 is inactive presumably because part of its reactive centre loop is inaccessible or does not have the conformation to bind its cognate proteases. The residues expected to interact with protease¹⁵, Ala 357–Glu 362 (P2–P4'), all reside in the extended loop on the surface of the molecule. The conformation of this loop is quite different from the arched loops formed by the reactive centres of other serine protease inhibitors¹⁵. Furthermore, Ala 357 (P2) and Glu 362 (P4') would be expected to interact with tissue plasminogen activator (tPA) in the tPA–PAI-1 complex^{15–18} but are partially buried in latent PAI-1.

Although the structure is not known of active PAI-1 or of any other active serpin, the structure of the intact, non-inhibitory serpin homologue ovalbumin has been reported¹¹. Ovalbumin lacks strand 4A, and contains strand 1C in a conformation similar to that found in α_1 -PI. The remainder of the reactive centre loop is folded into a surface helix (Fig. 3b). Active serpins may resemble ovalbumin in having strand 1C but, unlike ovalbumin, they probably contain a short, partially inserted strand 4A (refs 19–21). In ovalbumin, Arg 345 (P14) may prevent insertion of strand 4A (refs 11, 12). In other serpins, it is likely that the complete insertion of exposed reactive centre loop residues into strand 4A, either spontaneously (latent PAI-1) or upon cleavage (α_1 -PI), drives the conformational change underlying inactivation.

PAI-1 is the only serpin known to undergo a transition to the latent, inactive conformation without cleavage under physiological conditions¹⁴. The transition may be facile because there are few stereochemical barriers to the insertion of strand 4A. Mutagenesis has shown that the sequence of the reactive centre loop itself (between P16–P2') is not critical²². Therefore, strand insertion may be hindered by residues outside the reactive centre loop in other serpins but not in PAI-1. For example, residues Asn 49, Asp 177, Ser 330, Pro 391 and Gln 393 in α_1 -PI occupy the same space held by residues in the extended loop in latent PAI-1 (assessed by superposition of latent PAI-1 and α_1 -PI). The transition to the latent form is prevented *in vivo* by the binding to vitronectin in plasma^{3,23}.

It has been proposed that active serpins resemble ovalbumin¹¹. To make the transition from an ovalbumin-like structure to one resembling latent PAI-1, residues in the β ribbon 3C–4C (coloured red in Fig. 3b) would have to be displaced. Indeed, these residues (dashed in Fig. 2a) are disordered in latent PAI-1. Further, interactions among a cluster of hydrophobic residues (Phe 208, Pro 289, Pro 369, Phe 370 and Pro 391) which are invariant throughout the protein family tether this β ribbon to the rest of the molecule in other serpin structures⁷, but the interactions are absent in latent PAI-1. If active PAI-1 resembles ovalbumin in having a reactive centre loop positioned above sheet C (Fig. 3b), weakened interactions between the 3C–4C β ribbon and the rest of the protein could allow the collapse of the reactive centre loop to the position below sheet C seen in latent PAI-1.

The structural similarity of latent and cleaved PAI-1 was established by antibody cross-reactivity¹³, and was surmised from the similar thermostabilities of cleaved serpins^{5,24} and latent PAI-1 (refs 13, 14). Thermostability is likely to stem from insertion of strand 4A into sheet A, a structural feature common to the latent and cleaved molecules. Carrell *et al.*¹⁴ have demonstrated that other serpins can adopt a thermostable inactive state on incubation in denaturants at low concentration. We propose that these thermostable forms may closely resemble latent PAI-1 with respect to full insertion of strand 4A and the position and conformation of the extended surface loop. □

- Hekman, C. M. & Loskutoff, D. J. *J. Biol. Chem.* **260**, 11581–11587 (1985).
- Andreasen, P. A., Georg, B., Lund, L. R., Riccio, A. & Stacey, S. N. *Molec. cell. Endocr.* **68**, 1–19 (1990).
- Huber, R. & Carrell, R. W. *Biochemistry* **28**, 8951–8966 (1989).
- Bruch, M., Weiss, V. & Engel, J. *J. Biol. Chem.* **263**, 16626–16630 (1988).
- Schechter, I. & Berger, A. *Biochem. biophys. Res. Commun.* **27**, 157–162 (1967).
- Loebermann, H., Tokuda, R., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **177**, 531–556 (1984).
- Baumann, U. *et al. J. molec. Biol.* **218**, 595–606 (1991).
- Mourey, L. *et al. Biochimie* **72**, 599–608 (1990).
- Katagiri, K., Okada, K., Hattori, H. & Yano, M. *Eur. J. Biochem.* **176**, 81–87 (1988).
- Stein, P. E. *et al. Nature* **347**, 99–102 (1990).
- Wright, H. T., Qian, H. X. & Huber, R. *J. molec. Biol.* **213**, 513–528 (1990).
- Munch, M., Heegaard, C., Jensen, P. H. & Andreasen, P. A. *EMBO J.* (in the press).
- Carrell, R. W., Evans, D. L. & Stein, P. E. *Nature* **353**, 576–578 (1991).
- Papamokos, E. *et al. J. molec. Biol.* **158**, 515–537 (1982).
- Madison, E. L., Goldsmith, E. J., Gerard, R., Gething, M. J. & Sambrook, J. F. *Nature* **339**, 721–723 (1989).
- Madison, E. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 3530–3533 (1990).
- Madison, E. L., Goldsmith, E. J., Gething, M. J., Sambrook, J. F. & Gerard, R. D. *J. Biol. Chem.* **265**, 21423–21426 (1990).
- Kress, L. F., Kurecki, T., Chan, S. K. & Laskowski, M. *J. Biol. Chem.* **254**, 5317–5320 (1979).
- Schulze, A. J. *et al. Eur. J. Biochem.* **194**, 51–56 (1990).
- Skriver, K. *et al. J. Biol. Chem.* **266**, 9216–9221 (1991).
- Lawrence, D. A., Strandberg, L., Ericson, J. & Ny, T. *J. Biol. Chem.* **265**, 20293–20301 (1990).
- Wiman, B., Almquist, Å., Sigurdardottir, O. & Lindahl, T. *FEBS Lett.* **242**, 125–128 (1988).
- Carrell, R. W. & Owen, M. C. *Nature* **317**, 730–732 (1985).
- Jones, T. A. & Kjeldgaard, M. *O Version 5.4* (Dept Molec. Biol., Univ. Uppsala, Sweden, 1991).
- Franke, A. E. *et al. Biochim. biophys. Acta* **1037**, 16–23 (1990).
- Goldsmith, E. J. *et al. Proteins* **9**, 225–227 (1991).
- Howard, A. J., Nielsen, C. & Xuong, N.-H. *Meth. Enzym.* **114a**, 452–472 (1985).
- Crowther, R. A. *Acta crystallogr.* **B25**, 2571–2580 (1969).
- Crowther, R. A. & Blow, D. M. *Acta crystallogr.* **23**, 544–548 (1967).
- Fitzgerald, P. F. *J. appl. Crystallogr.* **21**, 273–278 (1988).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Carson, M. & Bugg, C. E. *J. molec. Graph.* **4**, 121–122 (1986).

ACKNOWLEDGEMENTS. We thank S. Sprang for helpful suggestions. This research was supported by the NIH and the American Heart Association.

MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis

Hisaji Maki & Mutsuo Sekiguchi

Department of Biochemistry, Faculty of Medicine, Kyushu University, Fukuoka 812, Japan

ERRORS in the replication of DNA are a major source of spontaneous mutations, and a number of cellular functions are involved in correction of these errors to keep the frequency of spontaneous mutations very low¹. We report here a novel mechanism which prevents replicational errors by degrading a potent mutagenic substrate for DNA synthesis. This error-avoiding process is catalysed by a protein encoded by the *mutT* gene of *Escherichia coli*, mutations of which increase the occurrence of A · T → C · G transversions 100 to 10,000 times the level of the wild type². Spontaneous oxidation of dGTP forms 8-oxo-7,8-dihydro-2'-dGTP (8-oxodGTP), which is inserted opposite dA and dC residues of template DNA with almost equal efficiency, and the MutT protein specifically degrades 8-oxodGTP to the monophosphate. This indicates that elimination from the nucleotide pool of the oxidized form of guanine nucleotide is important for the high fidelity of DNA synthesis.

We have shown that a small amount of labelled dGMP was incorporated onto poly(dA)/oligo(dT)₂₀ template by the action of *E. coli* DNA polymerase III and the misincorporated nucleotide was not eliminated by the editing function of the polymerase³. But, the misincorporation was specifically prevented *in vitro* by the MutT protein which hydrolyses nucleoside triphosphates with an apparent preference for dGTP^{3,4}. As the suppression of misincorporation occurred when a large amount of dGTP was present in the *in vitro* DNA synthesis system, we considered that the MutT protein might selectively degrade a mutagenic form of dGTP to prevent the dG·dA mispairing³. When dNMPs, prepared by P1 nuclease digestion from the DNA containing the misincorporated nucleotide, were subjected to

TABLE 1 Incorporation of 8-oxodGMP opposite dA and dC residues of template DNA

Template	Substrate	K_m (μ M)	V_{max}	V_{max}/K_m (M^{-1})	Relative efficiency
dA	dTTP	1.0	2.0	2.0×10^{-6}	1.0
dA	8-oxodGTP	11	0.85	7.7×10^{-8}	0.039
dC	dGTP	1.0	2.1	2.1×10^{-6}	1.0
dC	8-oxodGTP	5.3	0.32	6.0×10^{-8}	0.029

The relative velocity for incorporation of nucleoside monophosphate opposite a single residue of dA or dC of each template DNA was determined as described⁶. The data were collected from experiments with various concentrations of substrate nucleotides (0.5–5.0 μ M for dTTP and dGTP, and 5.0–50 μ M for 8-oxodGTP). Intensities of bands in autoradiograms were measured using a Fujix 2000 Bio Image Analyzer. K_m and relative V_{max} were obtained from Lineweaver–Burk plots¹⁷ of the data. Relative efficiency was calculated by dividing the V_{max}/K_m for incorporation of 8-oxodGMP by that of a correct substrate into each template DNA.

thin-layer chromatography, all the radioactivity associated with the misincorporated nucleotide migrated slower than the authentic dGMP (H.M., M. Akiyama, J.-Y. Mo and M.S., unpublished data). This meant that the misincorporated nucleotide is not dGMP but rather its derivative which is more negatively charged than dGMP. As deoxynucleotide carrying 8-oxoG has such properties^{5,6}, we suspected that 8-oxodGTP, present as a minor component in the labelled dGTP preparation that we used, may be inserted opposite dA residues of the template. The observation that dCMP and dAMP were selectively and equally incorporated opposite the 8-oxodG residue of a synthetic DNA template⁷ also supported this view. All these findings prompted us to investigate whether 8-oxodGMP can be incorporated into DNA.

Using the α subunit of *E. coli* DNA polymerase III, which has DNA polymerizing capacity but not 3' $\rightarrow$ 5' exonuclease activity⁸, we examined the incorporation of 8-oxodGMP onto synthetic DNA templates (Fig. 1). 8-oxodGTP was efficiently inserted opposite the dA residue as well as dC on the template.

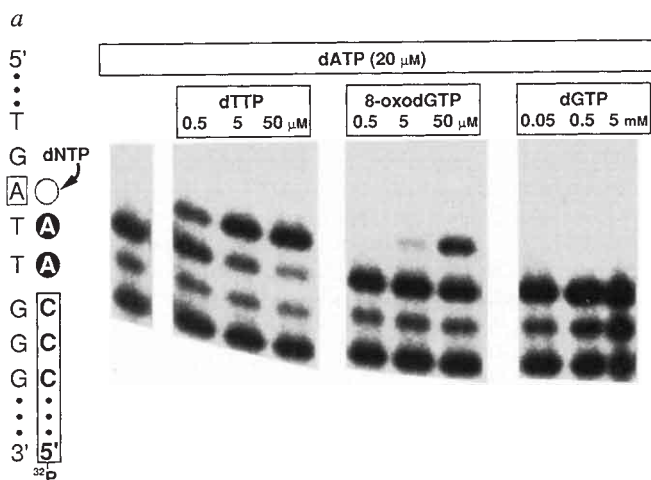


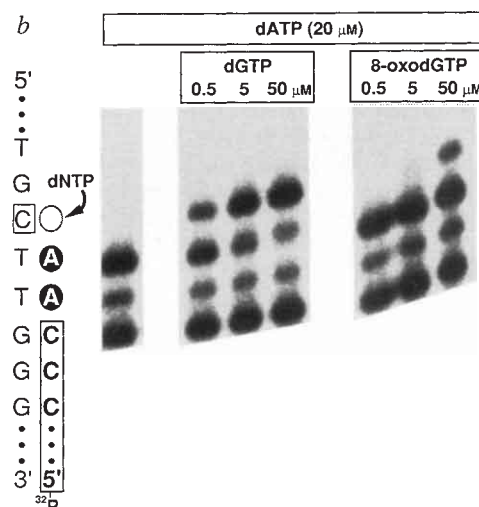
FIG. 1 Incorporation of 8-oxodGMP opposite dA and dC residues of template DNA. 8-oxodGTP was prepared by the method of Kasai and Nishimura⁶ with slight modification and purified by two successive column chromatography steps with Spherisorb SAX and DEAE-MemSep (details to be published elsewhere). Purity and quantity of the preparation were examined in electrochemical measurements and ultraviolet spectrum analyses. dATP, dTTP and dGTP were from Pharmacia-LKB (FPLC-pure grade). Oligodeoxynucleotides were prepared by an automated DNA synthesizer (ABI model 381A) and purified by reversed-phase column chromatography. dA- and dC-template DNAs (3'-CGTCGGTTTTGCAGGGTT^A/cGTTAGTATG-5') differ only in the tenth nucleotide from their 5'-termini. 5'-end labelling of primer DNA (5'-GCAGCCAAACGTC-3') and annealing of the primer to template DNAs

TABLE 2 Substrate specificity of MutT nucleoside triphosphatase

Substrate	K_m (μ M)	V_{max}	V_{max}/K_m
8-oxodGTP	0.48	4.2	8,800
dGTP	1,100	4.8	4.3
dTTP	1,800	1.3	0.72
dCTP	1,700	0.28	0.16
GTP	1,200	1.6	1.4

Rates of hydrolysis of nucleoside triphosphates were determined by time-course experiments with eight different concentrations of the substrates (0.05–4.1 μ M for 8-oxodGTP and 0.2–4.0 mM for other nucleotides). [α -³²P]dNTPs and -GTP were from Amersham. Unlabelled nucleotides were FPLC grade (Pharmacia-LKB), except for 8-oxodGTP. Hydrolysis of nucleoside triphosphates by the MutT protein was as described in the legend to Fig. 2. Amounts of the products were measured using a Fujix 2000 Bio Image Analyzer. Knowing the relative molecular mass of MutT protein¹⁸ (15,000) and assuming that the preparation of protein was fully active, the turnover number (molecules of nucleoside monophosphate produced by one molecule of MutT protein in a second) was calculated. From Lineweaver–Burk plots¹⁷ of the data, K_m and V_{max} for hydrolysis of each nucleoside triphosphate were obtained.

There was no evidence for incorporation of 8-oxodGMP opposite dG and dT (data not shown). The Michaelis constant (K_m) and the maximum velocity (V_{max}) of each reaction are shown in Table 1. The apparent K_m for incorporation of 8-oxodGMP opposite dA and dC is only 5 to 10 times greater than those for the normal pairing (dTTP opposite dA and dGMP opposite dC). Relative efficiencies of incorporation of 8-oxodGMP opposite dA and dC were almost the same, and these values were 100- to 1,000-fold higher than those for misincorporation of usual nucleotides by the α subunit (misincorporation of dCMP or dAMP opposite dA and that of dGMP opposite dT)⁹. Note that misincorporation of dGMP opposite dA residue was not evident when a highly purified preparation of dGTP was used for the reaction (Fig. 1a). Thus, it is likely that previous observations^{3,9} of *in vitro* misincorporation of



were as described⁹. DNA synthesis on the dA-template was done with 1.4 units of highly purified preparation of α subunit⁸ at 30 °C for 10 min. Reaction mixtures contained 20 μ M dATP and varying concentrations of dTTP, 8-oxodGTP or dGTP. Other procedures were as described⁹. Autoradiograms were obtained using a Fujix 2000 Bio Image Analyzer (Fuji). a, Gel autoradiogram showing the incorporation of 8-oxodGMP opposite dA residues of template DNA. b, Gel autoradiogram showing the incorporation of 8-oxodGMP opposite dC residues of template DNA.

one functioning at substrate level (for the MutT protein) and the other at the DNA level (for the MutM (Fpg) protein). □

Received 11 September; accepted 14 October 1991.

1. Horiuchi, T., Maki, H. & Sekiguchi, M. *Bull. Inst. Pasteur* **87**, 309–336 (1989).
2. Yanofsky, C., Cox, E. C. & Horn, V. *Proc. natn. Acad. Sci. U.S.A.* **55**, 274–281 (1966).
3. Akiyama, M., Maki, H., Sekiguchi, M. & Horiuchi, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3949–3952 (1989).
4. Bhatnagar, S. K. & Bessman, M. *J. Biol. Chem.* **263**, 8953–8957 (1988).
5. Kasai, H. & Nishimura, S. *Nucleic Acids Res.* **12**, 2137–2145 (1984).
6. Kasai, H. & Nishimura, S. in *Oxidative Stress: Oxidants and Antioxidants* (ed. Sies, H.), 99–116 (Academic, London, 1991).
7. Shibutani, S., Takeshita, M. & Grollman, A. P. *Nature* **349**, 431–434 (1991).
8. Maki, H. & Kornberg, A. *J. Biol. Chem.* **260**, 12987–12992 (1985).
9. Sloane, D. L., Goodman, M. F. & Echols, H. *Nucleic Acids Res.* **16**, 6465–6475 (1988).
10. Kasai, H., Tanooka, H. & Nishimura, S. *Gann* **75**, 1037–1039 (1984).
11. Wood, M. L., Dizdaroğlu, M., Gajewski, E. & Essigman, J. M. *Biochemistry* **29**, 7024–7032 (1990).
12. Moriya, M. *et al. Mutation Res.* **254**, 281–288 (1991).
13. Tchou, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 4690–4694 (1991).
14. Cabrera, M., Nghiem, Y. & Miller, J. H. *J. Bact.* **170**, 5405–5407 (1988).
15. Michaels, M. L., Pham, L., Cruz, C. & Miller, J. *Nucleic Acids Res.* **19**, 3629–3632 (1988).
16. Boosalis, M. S., Petruska, J. & Goodman, M. F. *J. Biol. Chem.* **262**, 14689–14696 (1987).
17. Fersht, A. in *Enzyme Structure and Mechanism* 91–92 (Freeman, San Francisco, 1977).
18. Akiyama, M., Horiuchi, T. & Sekiguchi, M. *Molec. gen. Genet.* **206**, 9–16 (1987).
19. Bradford, M. M. *Analyt. Biochem.* **72**, 248–254 (1976).

ACKNOWLEDGEMENTS. We thank M. Ohara for comments. This work was supported by the Ministry of Education, Science and Culture of Japan.

Identification of a Fab interaction footprint site on an icosahedral virus by cryoelectron microscopy and X-ray crystallography

Guoji Wang, Claudine Porta, Zhongguo Chen, Timothy S. Baker & John E. Johnson*

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

BIOLOGICAL processes frequently require the formation of multi-protein or nucleoprotein complexes. Some of these complexes have been produced in homogeneous form, crystallized, and analysed at high resolution by X-ray crystallography (for example, see refs 1–3). Most, however, are too large or too unstable to crystallize. Individual components of such complexes can often be purified and analysed by crystallography. Here we report how the coordinated application of cryoelectron microscopy, three-dimensional image reconstruction, and X-ray crystallography provides a powerful approach to study large, unstable macromolecular complexes. Three-dimensional reconstructions of native cowpea mosaic virus (CPMV) and a complex of CPMV saturated with a Fab fragment of a monoclonal antibody against the virus have been determined at 23 Å resolution from low-irradiation images of unstained, frozen-hydrated samples. Despite the nominal resolution of the complex, the physical footprint of the Fab on the capsid surface and the orientation and position of the Fab have been determined to within a few ångströms by fitting atomic models of CPMV⁴ and Fab (Kol)⁵ to reconstructed density maps.

Comoviruses are plant viruses whose structural and biological properties are strikingly similar to the animal picornaviruses^{6,7}. For example, the protein capsids of CPMV, the type member of the comovirus group, and poliovirus, the most extensively studied picornavirus, are comparable (Fig. 1a). The CPMV coat contains 60 copies each of large (relative molecular mass 42,000 (*M_r* 42K) and small (24K) protein subunits arranged with icosahedral symmetry. Each large subunit (L) is composed of two antiparallel β-barrel domains which are positioned close to the icosahedral 3-fold axes. The β-barrel domain at the

* To whom correspondence should be addressed.

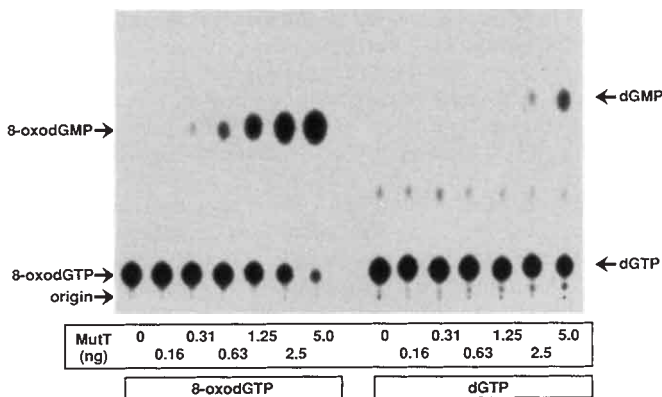


FIG. 2 Hydrolysis of 8-oxodGTP and dGTP by MutT protein. [α -³²P]8-oxodGTP was prepared from [α -³²P]dGTP (Amersham) and purified. [α -³²P]dGTP from Amersham was diluted with unlabelled dGTP (FPLC-pure grade, Pharmacia-LKB) and used without further purification. Protein concentration of a homogeneous preparation of MutT protein³ was determined by the method of Bradford¹⁹ with BSA as a standard. Hydrolysis reactions were at 30 °C for 10 min with varying amounts of MutT protein. The reaction mixture (12.5 μ l) contained 40 μ M 8-oxodGTP or dGTP, and other conditions were as described³. The production of 8-oxodGMP or dGMP was followed by thin-layer chromatography (PEI-cellulose plate, Merck) with 1 M LiCl and the autoradiogram was processed using a Fujix 2000 Bio Image Analyzer. Reaction mixtures (1 μ l) were spotted on the same plate.

dGMP opposite dA were due to a minute amount of 8-oxodGTP present in the crude preparations of dGTP.

To examine whether the MutT protein efficiently hydrolyses 8-oxodGTP, we titrated the protein in a nucleoside triphosphatase assay with 40 μ M 8-oxodGTP or dGTP and analysed the production of nucleoside monophosphates by thin-layer chromatography (Fig. 2). 8-oxodGTP was degraded 10 times faster than dGTP under the conditions used. This seemed to be due to a strong affinity of the protein to 8-oxodGTP. The kinetic parameters for hydrolysis of several nucleoside triphosphates by the MutT protein were measured (Table 2). The apparent *K_m* for the hydrolysis of 8-oxodGTP was 2,000 times lower than that of dGTP, whereas the *V_{max}* for both nucleotides was the same. The MutT protein hydrolysed other dNTPs and GTP with lower *V_{max}* and extremely high *K_m* values. Thus we can conclude that 8-oxodGTP is a specific substrate for the MutT protein.

8-OxodG was first described as a major lesion of DNA after exposure to reducing agents⁵. This lesion is caused by active oxygen species produced by many naturally occurring compounds, including cellular metabolic intermediates⁶, as well as X-ray irradiation¹⁰. 8-oxodG is biologically important because a G·C → T·A transversion was induced at the site when DNA containing the damaged guanine was introduced into *E. coli* cells^{11,12}. Results consistent with this were obtained in an *in vitro* DNA synthesis study using a synthetic DNA template containing 8-oxodG (ref. 7). An enzyme that specifically removes 8-oxoguanine from DNA has been detected in *E. coli*¹³. As inactivation of this enzyme by *mutM* (*fpg*) mutation leads to a 10-fold increase of G·C → T·A transversion frequency^{14,15}, the oxidation of the C-8 position of guanine residues of DNA would seem to occur spontaneously in *E. coli* cells.

As the oxidation of guanine proceeds *in vitro* more rapidly in dGTP than in DNA (unpublished data), it is conceivable that the oxidative damage to the base would occur more frequently in the nucleotide pool of *E. coli* cells than in their chromosomal DNA. From the distinct mutator action of *mutT*, we propose that the MutT protein has a major role in preventing the A·T → C·G transversion caused by 8-oxodGTP, formed spontaneously in the nucleotide pool. To protect genetic information from oxidation damage, cells seem to have at least two mechanisms,

amino-terminal end of L corresponds to the VP2 subunit in poliovirus and the β -barrel at the carboxy-terminal end of L corresponds to VP3 in poliovirus. The small subunits (S), packed near the icosahedral 5-fold axes, correspond to VP1 in poliovirus.

CPMV is strongly immunogenic when injected into rabbits or mice. We prepared nine monoclonal antibodies against CPMV using standard methods⁸. The binding properties of these antibodies were analysed by enzyme-linked immunosorbent assay (ELISA) and all nine would only bind to virus particles presented in a double antibody sandwich-type ELISA in which the virus is bound to a polyclonal antibody attached to the plastic plate. Denaturation of the virus by its direct adsorption onto a plastic surface destroys the antigenic sites for all nine monoclonal antibodies. Only monoclonal antibodies 5B2 and 10B7 reacted strongly with CPMV that was free in solution (antibody binding was assayed by electron microscopy analysis of negatively stained complexes marked with gold-labelled anti-mouse immunoglobulin). Fab fragments from monoclonal antibody 5B2 were prepared by digestion of IgGs with insoluble papain (Sigma). The papain was removed by low-speed centrifugation and the digestion products were incubated with CPMV for 2 h at 37 °C before centrifugation of the reaction mixture on a 10–40% sucrose gradient to separate Fab and virus/Fab complexes. A single peak containing virus and virus complexed with varying amounts of Fab was fractionated and visualized by electron microscopy. Unstained aliquots of native CPMV and of virus saturated with Fab were quick-frozen and low-irradiation images (Fig. 1b, c) were recorded to compute three-dimensional reconstructions (Fig. 1d, e).

The α -backbone model of CPMV, derived from the 3.0 Å-resolution X-ray structure⁴, was placed in both enantiomorphs of the native (Fig. 1d) and the virus-Fab-complex (Fig. 1e) reconstructed densities (the correct enantiomorph of the electron microscope density cannot be determined from a single micrograph). The program FRODO⁹ was used to align the symmetry

axes of the X-ray and electron microscope structures and to compute an overall scale factor to adjust for small errors in the microscope magnification. The exceptionally good fits of the X-ray model with both reconstructions allowed the correct enantiomorph of the density to be unambiguously identified and suggests that Fab binding induces no major change in the capsid structure.

The Fab density, centred 25 Å from the icosahedral 3-fold symmetry axes, covers an ellipsoidal area roughly 40×20 Å on the viral surface and extends radially about 70 Å from it (Figs 1e and 2b). The overall shape and extent of the virus-Fab interaction agrees very well with protein-Fab interactions that have been analysed at high resolution by X-ray crystallography¹. The roughly 30 amino-acid residues of the virus that lie under the 'footprint' of the Fab were identified from inspection of the high-resolution X-ray model of CPMV (Fig. 2c). The Fab footprint covers regions of the CPMV surface that correspond to subunits VP2 and VP3 in poliovirus. This antigenic region is composed of a quaternary interaction between the 'VP2' β -barrel domain of one L subunit and the 'VP3' β -barrel domain in a neighbouring 3-fold-related L subunit. Exposed loops from each of these domains are in favourable positions to interact with the Fab and each loop contains a residue spatially equivalent to the corresponding two amino acids that define the 3B antigenic site of poliovirus¹⁰ (mutation of either residue in poliovirus prevents neutralization by a specific monoclonal antibody to poliovirus) (Fig. 2d). The common immunogenic sites on CPMV and poliovirus probably arise because of the overall similarity in the capsids of the two viruses. Some immunogenic regions on mammalian viruses are readily mutable as a result of evolutionary pressure exerted by the circulating immune system on exposed elements of the structure. Other sites may be surface regions that are good immunogens, but lack the structural flexibility to readily mutate. Page *et al.*¹⁰ point out that immunogenic site 3B in type I poliovirus is likely to be such an epitope because of the small number of viable mutants observed. This prediction

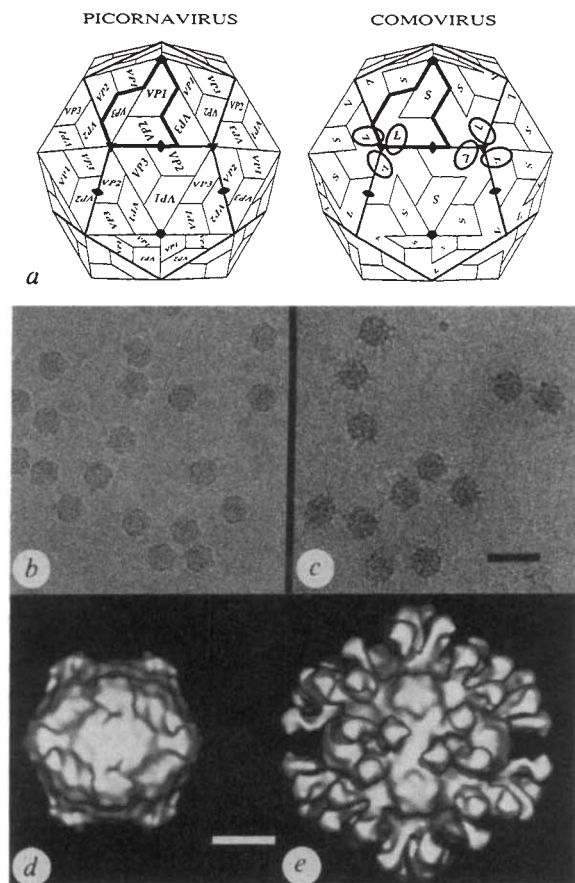
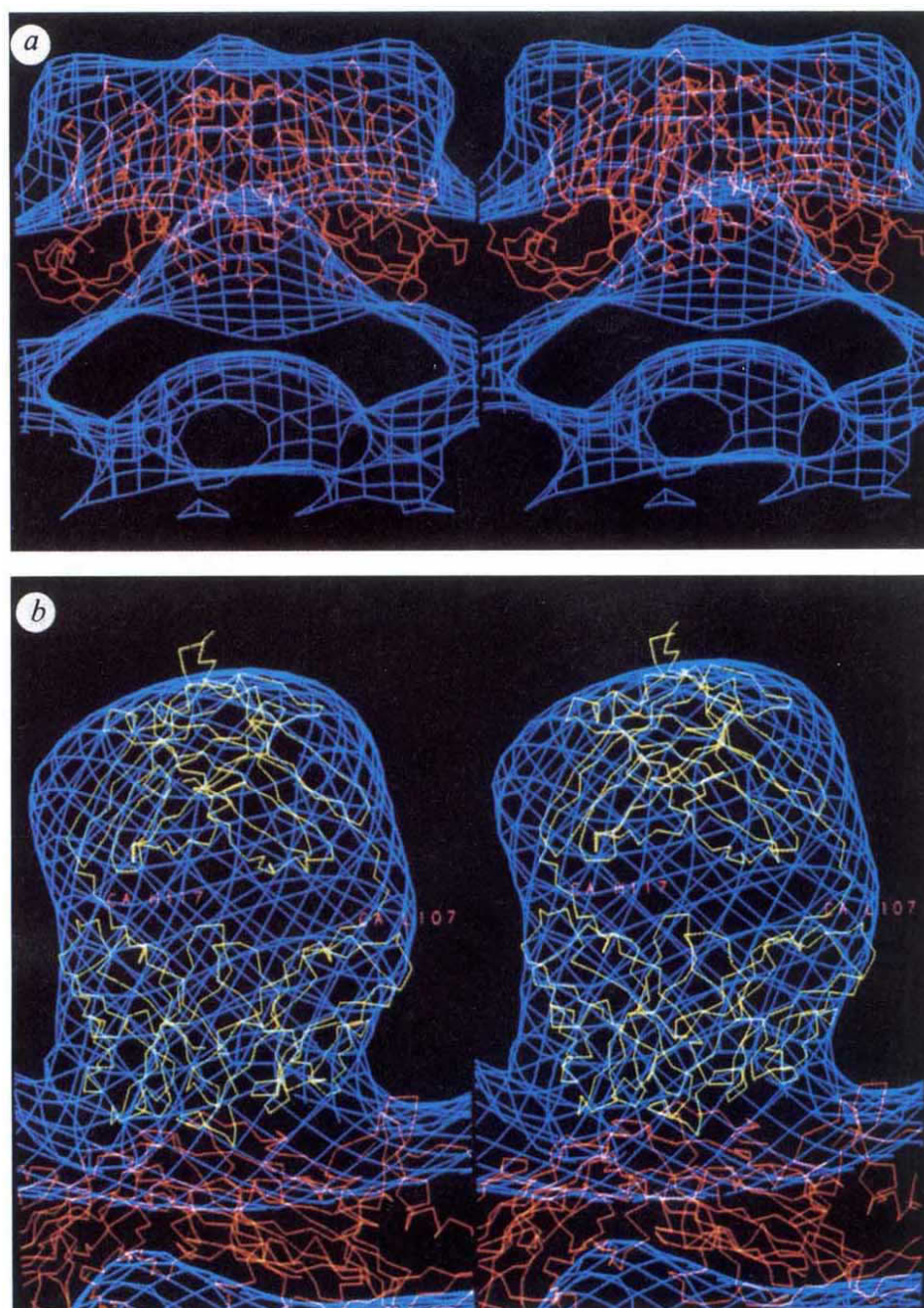


FIG. 1 The capsid architectures of picorna and comoviruses displayed with electron micrographs and three-dimensional image reconstructions of frozen-hydrated native CPMV and CPMV-Fab complexes. a, Schematic comparison of picornavirus and comovirus capsids. In each case one trapezoid represents an eight-stranded antiparallel β -barrel (see Fig. 2d). The icosahedral asymmetric unit of the picornavirus capsid (shaded in the heavy outline) contains three β -barrels, labelled VP1, VP2 and VP3, each with a characteristic amino-acid sequence. The comovirus capsid is similar to the picornavirus capsid except that two of the β -barrels, corresponding to VP2 (N terminus) and VP3 (C terminus), are covalently linked to form a single polypeptide, the large protein subunit (L), and the small protein subunit (S) corresponds to VP1. The ellipsoids on the comovirus capsid show the rough location of 6 of the 60 equivalent Fab binding sites (compare with e). b and c, Electron micrographs of frozen-hydrated samples of CPMV (b) and CPMV complexed with monoclonal antibody (5B2) Fab fragments (c). Scale bar, 500 Å. Both samples were prepared and photographed with established cryomicroscopy procedures^{14–17}. Roughly 4 ml each of a 2.0 mg ml^{-1} CPMV (b) or 0.5 mg ml^{-1} CPMV-Fab (c; stoichiometry of Fabs to virus in solution was 600:1) sample at pH 7.4 in PBS was applied to holey-carbon films, blotted with filter paper and rapidly plunged into liquid ethane at about -170 °C. Images were recorded at minimal electron dose ($\sim 16 \text{ e}^- \text{ Å}^{-2}$) at a nominal magnification of $\times 49,000$ and at 80 kV in a Philips EM420 electron microscope. The images in b and c were recorded at about 0.8 and 0.9 μm underfocus, respectively. d, e, Surface-shaded representations of three-dimensional reconstructions of native virus (d) and virus-Fab complex (e), both viewed in the standard crystallographic 2-fold orientation, were computed with established methods^{18–20}. A total of 17 (d) or 22 (e) particle images were combined to reconstruct both structures to a resolution limit of 23 Å. Scale bar, 100 Å.

FIG. 2 Comparisons of CPMV and Fab atomic models with reconstructed electron density derived from electron micrographs of frozen-hydrated specimens. *a*, *b*, Stereoviews comparing α backbone models of CPMV (red) and Fab (Kol) (yellow) with electron density (blue) derived from the electron-microscopy reconstruction described in Fig. 1. *a*, Native virions viewed in a direction perpendicular to a vertical fivefold icosahedral symmetry axis. The surface bulge in the density corresponds to amino acids in a β -turn between the outermost strands of the S subunit. *b*, A close-up view showing the Fab (Kol) and CPMV α backbones fitted to the CPMV-Fab reconstruction. The small portions of Kol that protrude outside the Fab electron microscope envelope at the top and into the surface of the CPMV L subunit are regions of the polypeptide that are not conserved in sequences of other IgG molecules, and are therefore likely to be different in the 5B2 Fab structure. When contoured at a higher density level the Fab envelope divides into well-defined globular regions which correspond closely with the Fab two-domain structure. Note that the electron microscope envelope, contoured at a single density level, fits the CPMV and Fab X-ray models equally well. This suggests that the frozen-hydrated CPMV samples imaged were nearly fully saturated with the Fabs and most of the Fab molecules must be fairly rigidly attached in equivalent orientations with respect to the viral surface. ▶



is consistent with the discovery of a similar epitope on CPMV. By contrast, site 2 on type I poliovirus displays a large number of viable mutations positioned on exposed loops that are not present in CPMV.

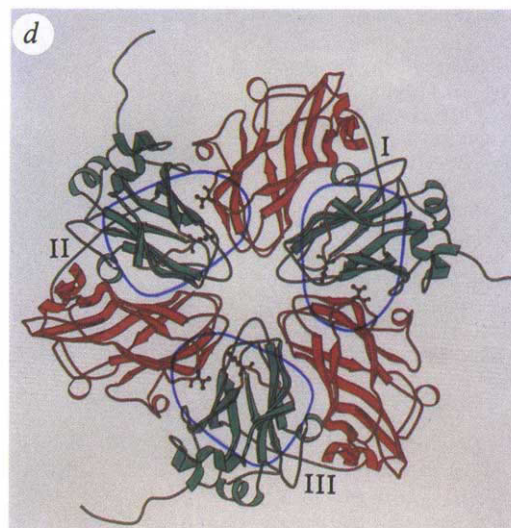
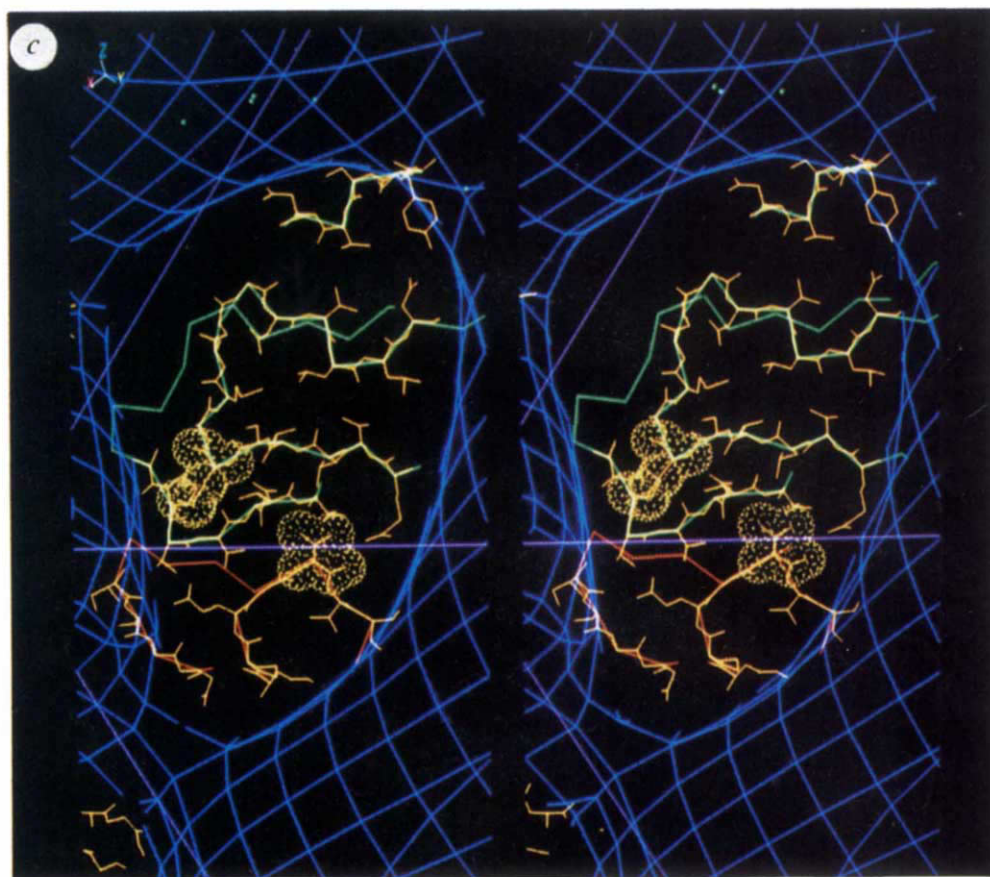
An atomic structure for Fab 5B2 is unknown, although atomic coordinates for other Fab fragments are available in the Brookhaven protein data bank. The α coordinates for Fab (Kol)⁵ were used for initial modelling of the relevant density in the CPMV-Fab reconstruction because Fab (Kol) has a 165° elbow angle (the angle between pseudo 2-fold axes relating heavy and light chains in the variable domain and heavy and light chains in the constant domain), which is consistent with the extended shape of the density ascribed to the Fab in the reconstruction. The variable domain of the Fab (Kol) model was positioned in contact with the virus surface and adjusted as a rigid body to fit the electron microscope density. Only two orientations of the

Fab, which were related by a 180° rotation about the axis perpendicular to the particle surface, gave excellent agreement between the model and the ellipsoidal-shaped density.

The factors determining virus neutralization by antibodies are not fully understood. But for some antibodies, bivalent binding of the IgG was determined to be important to its ability to neutralize the virion¹¹. The centres of pairs of 5B2 Fab fragments, related by an adjacent icosahedral 2-fold symmetry axis, are 60 Å apart. This is near the minimum distance thought to be reasonable for bidentate binding of an IgG¹². Model building shows that bivalent binding of the 5B2 IgG is possible if its Fab portions can adopt a smaller elbow angle.

The excellent correspondence of the X-ray and electron microscope structures for both the Fab and CPMV density distributions demonstrate the high fidelity of reconstructed density maps computed from images of frozen-hydrated specimens. Even

c, Stereoview of the Fab footprint on the virion surface showing regions of the CPMV polypeptide chain that lie under the Fab. The region in red corresponds to the 'VP3' domain of one L subunit; green denotes the 'VP2' domain of a 3-fold-related L subunit. Main chain and side chains of residues that may interact with the Fab are shown in yellow. The amino acids displayed with surface rendering (Lys 34 in the 'VP2' domain and Thr 24 in the 3-fold-related 'VP3' domain) are spatially equivalent to residues 72 and 76 in the poliovirus VP2 and VP3 domains. Mutations of either of these residues in poliovirus prevents neutralization by poliovirus monoclonal antibodies 10 and 13 (ref. 10), thus defining the 3B antigenic site on serotype 1 poliovirus. Lines in magenta correspond to the boundaries in Fig. 1a (picornavirus) that separate VP2 from VP3 near the 3-fold axis on the left side of the figure. d, A schematic representation²¹ of 3-fold-related L subunits, each numbered near the peptide that connects 'VP2' (green) and 'VP3' (red) domains. Ellipsoids identify the viral surface in contact with each Fab. The side chains shown correspond to the residues with surface rendering in Fig. 2c. The ellipsoid on the right (covering portions of L subunits I and III) is in the same orientation as the ellipsoid in c.



though it is impossible to accurately model the detailed interactions between the Fab and the virus with these data, our results clearly suggest candidates for site-directed mutagenesis of an infectious clone¹³ to explore the role of individual amino acids in complex formation. The methods used for the analysis of the CPMV-Fab complex can be extended to study other viruses or macromolecules and, for example, reveal details of critical molecular interactions that occur between them and antibodies or receptor molecules. □

Received 20 August; accepted 25 October 1991.

1. Davies, D. R., Padlan, E. A. & Sheriff, S. A. *Rev. Biochem.* **59**, 439-473 (1990).
2. Ruff, M. *et al. Science* **252**, 1682-1689 (1991).
3. Rossmann, M. G. & Johnson, J. E. *A. Rev. Biochem.* **58**, 533-573 (1989).
4. Chen, Z., Stauffer, C. V. & Johnson, J. E. *Semin. Virol.* **1**, 453-466 (1990).
5. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. *J. molec. Biol.* **141**, 369-391 (1980).
6. Goldbach, R. & van Kammen, A. in *Molecular Plant Virology* Vol. 2 (ed. Davies, J. W.) 83-120 (CRC, Boca Raton, Florida, 1985).
7. Chen, Z. *et al. Science* **245**, 154-159 (1989).
8. Fazekas de St. Groth, S. & Scheidegger, D. *J. immunol. Meth.* **35**, 1-121 (1980).
9. Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 303-317 (Clarendon, Oxford, 1982).
10. Page, G. S. *et al. J. Virol.* **62**, 1781-1794 (1988).
11. Icenogle, J. H. *et al. Virology* **127**, 412-425 (1983).
12. Mosser, A. G., Leippe, D. M. & Rueckert, R. R. in *Molecular Aspects of Picornavirus Infection and Detection* (eds Semler, B. L. & Ehrenfeld, E.) 155-167 (Am. Soc. Microbiol., Washington, DC, 1989).

13. Holness, C. L., Lomonosoff, G. P., Evans, D. & Maule, A. J. *Virology* **172**, 311-320 (1989).
14. Adrian, M. J., Dubochet, J., Lepault, J. & McDowell, A. W. *Nature* **308**, 32-36 (1984).
15. Milligan, R. A., Brisson, A. & Unwin, P. N. T. *Ultramicroscopy* **13**, 1-10 (1984).
16. Dubochet, J. *et al. Q. Rev. Biophys.* **21**, 129-228 (1988).
17. Prasad, B. V. K., Burns, J. W., Marietta, E., Estes, M. K. & Chiu, W. *Nature* **343**, 476-479 (1990).
18. Baker, T. S., Drak, J. & Bina, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 422-426 (1988).
19. Baker, T. S., Newcomb, W. W., Booy, F. P., Brown, J. C. & Steven, A. C. *J. Virol.* **64**, 563-573 (1990).
20. Yeager, M., Dryden, K. A., Olson, N. H., Greenberg, H. B. & Baker, T. S. *J. Cell. Biol.* **110**, 2133-2144 (1990).
21. Kraulis, P. J. *J. Appl. Crystallogr.* (in the press).

ACKNOWLEDGEMENTS. G.W. and C.P. are joint first authors. We thank Y. Li for virus preparations, M. H. V. van Regenmortel for help with preparation of the monoclonal antibodies and for support, R. Holland Cheng for help in converting electron density maps for different display systems, S. Fateley for help in preparing the manuscript and D. L. D. Caspar for critically reading the manuscript. This work was supported by NIH (T.S.B. and J.E.J.), a National Science Foundation Grant (T.S.B.), and The Lucille P. Markey Trust.

Mixed micelles in drug delivery

Danilo D. Lasic

Large disk-like mixed micelles composed of a drug and biological lipid are thermodynamically stable and represent a novel drug delivery system. Their unique physical properties are reflected in a significantly improved therapeutic index.

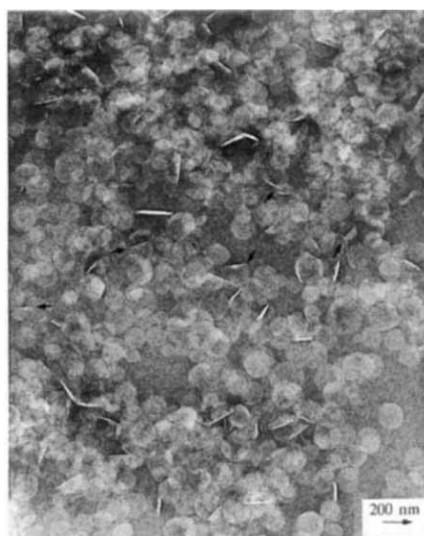
SMALL differences between therapeutic and toxic concentration severely limit the clinical utility of many potent drugs. In addition, drug molecules that cannot be absorbed upon oral administration and which have to be administered parenterally are often insoluble in water. In some cases, the therapeutic efficacy can be improved by delivering the drug with an appropriate microcarrier system¹, which is able to change temporal and spatial distribution (that is, pharmacokinetics and biodistribution), of the drug.

The problem of solubility can be solved by dissolving the drug in micelles². In general, this is a poor solution because micelles are very unstable upon dilution and interact with various blood components after intravenous administration. Blood circulation times and disassociation of drug from microcarriers can sometimes be improved by complex mixed micellar systems³, or (co)polymerized micelles of amphiphilic (pro)drugs⁴. Here we report on the physical, chemical, and biological properties of a new kind of drug delivery vehicle — large, thermodynamically stable, disk-like micelles of an antifungal drug and a biological lipid.

Drug and drug formulations

Amphotericin B, a lipophilic polyene antibiotic, is the drug of choice in the treatment of disseminated fungal infections occurring as a result of reduced immunocompetence (seen in chemotherapy and AIDS patients), and are frequently fatal. Unfortunately, the dosing of Amphotericin B is severely limited due to its toxic side effects, which include acute chills and fevers and chronic nephrotoxicity. In order to decrease these side effects and to circumvent the poor solubility of the drug in water, several drug carrier systems have been developed, ranging from deoxycholate micelles (Fungizone), liposomal preparations⁵ and ribbon-like lyotropic aggregates⁶ to disk-like colloidal particles⁷, which are the subject of this report.

The Amphotericin B colloidal dispersion (Amphocil) is prepared by isothermally injecting an equimolar DMSO solution of Amphotericin B and cholesterol sulphate at 50 °C into a buffer at pH 7.4. The DMSO is then removed by diafiltration. Cryoprotectant lactose is added before sterile filtration and lyophilization. The final concentration of Amphotericin B is 5 mg ml⁻¹ in 9.5% lactose.



A negative staining electron micrograph of the Amphotericin B colloidal dispersion. Most of the disk-like micelles lie flat on the support. The needle-like structures are micelles that are positioned perpendicularly, as can be inferred from the few micelles that folded during drying (see arrows). Bar indicates 200 nm (adapted from ref. 7 with permission from Elsevier).

The size distribution and shape of particles was determined by several spectroscopic and diffraction techniques⁷ and an electron micrograph is shown in the figure. The Amphocil particles show exceptional stability: in liquid form the size distribution does not change after several months of storage at 30 °C. The formulation can be lyophilized and reconstituted without any detectable changes. Such formulations are physically and chemically stable for more than two years when stored at 30 °C. In addition, their size does not change upon dilution, nor do they react with plasma proteins and blood cells⁷.

Most micelles are small, spherical structures which, at higher concentrations, ionic strengths, or temperatures, may aggregate ('cloud point') or grow into rod-like or disk-like micelles with asymmetry parameters a/b normally not larger than 3. Even disk-like micelles in the liquid-crystalline phase do not show large asymmetries because large deviations from the spheroidal shapes are normally energetically and entropically unfavoured.

These large, disk-like micelles may represent an important and novel structure, which can be theoretically predicted

from a model of vesicle formation⁸. According to this model, a large disk-like micelle is an intermediate structure in the vesicle formation process for methods in which the lipid molecules are dispersed at random orientations in micelles, or (im)miscible organic solvents⁹. For instance, in the liposome preparation method by detergent depletion small detergent-lipid mixed micelles fuse into disk-like structures. These micelles resemble small fragments of lipid bilayer with detergent molecules shielding the unfavourable exposure of hydrophobic parts of lipid molecules against water at their edges. When the radii of these micelles grow above some critical value and when there is not enough detergent for effective shielding they bend and self-close into vesicles.

The two energy contributions that dictate this closure are the interaction at the edge of the micelles (E) and curvature elasticity (C). E is proportional to the disk's circumference by energy per unit length (γ) and C is proportional to the disk area and inversely proportional to the square of the curvature radius. For closed surfaces, C is constant and equals $8\pi K$, with K being a curvature bending elasticity.

Thermodynamic stability

In most cases, these large disk-like micelles are not stable and without an energy input, which causes vesicle formation, they would aggregate/flocculate. However, two other scenarios are possible: very low value of C (that is, γ) and very low value of E (that is, K).

The first case leads to spontaneous vesiculation¹⁰, the second to large, thermodynamically stable, disk-like micelles described in this report. In both cases, the structures are entropically stabilized, that is, $8\pi K \sim kT$ (Boltzmann constant and temperature, respectively) for spontaneous vesiculation and $2\pi\gamma \sim kT$ for stable micelles with radius r . For $r=70$ nm, one gets $\gamma \sim 9.2 \cdot 10^{-15}$ J m⁻¹, while for spontaneous vesiculation $K \leq 1.6 \cdot 10^{-22}$ J. The low value of γ can be understood by the chemical structure of the Amphotericin B molecule, which has nine hydroxyl groups on one side while the other side of the rectangle-like molecule is hydrophobic. By proper packing, alignment and orientation of the Amphotericin B and cholesterol sulphate molecules, these hydroxyl groups can form almost a 'nonin-

teractive', ice-like structure with surrounding water molecules.

The problem of defining thermodynamic stability is not an easy one. The situation is especially unclear when dealing with colloids. Chemically identical systems, including Amphocil, can be more or less stable in several different physical states, which depend on preparation pathway and its kinetics. In many cases, physical stability cannot even be tested because of constituent chemical instability. In general, with amphiphilic systems one can use the term 'thermodynamically stable' for lyotropic liquid crystals but not particulate dispersions. Amphocil can be described as 'thermodynamically stable' based on the widely accepted assumption among researchers in the field that the term be used to describe systems that are stable for time periods of at least a few weeks to months. In addition, the above theoretical analysis shows that the particles are entropically stabilized.

Pharmacology

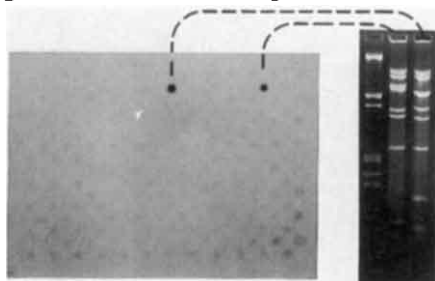
The physical properties of Amphocil are reflected in the pharmacology of this dosage form^{7,11,12}. Although Amphocil appears to possess the same spectrum of pharmacological and toxicological activities as deoxycholate micelles of Amphotericin B, it offers a significant improvement in therapeutic index in animal studies. Acute and repeat-dose toxicology studies in mice, rats, and dogs have shown that Amphocil is 5–15-fold safer than deoxycholate micelles of Amphotericin B⁷. When administered to mice infected with *Candida* or *Cryptococcus*, Amphocil is equipotent compared to micellar Amphotericin B, whereas against coccidioidomycosis, Amphocil is approximately half as potent⁷. Thus, the therapeutic index of Amphocil is improved 4–6 fold over micellar Amphotericin B⁷.

Although the origin of the improved biological activity is not yet known, one may speculate that it is due to the unique physical properties of the Amphocil complex as compared to micellar, liposomal, or liquid-crystalline formulations: tight drug binding in the complex, small size, and extremely high drug-to-lipid weight ratio, all of which improve the biodistribution and pharmacokinetics of the drug. Despite the fact that this drug delivery vehicle is clearly not as universally applicable as liposomes or microemulsions, several drugs may form similar structures at appropriate molar ratios with various amphiphilic lipids or steroids. More recently, diacyl-lipids with polymeric polar heads were introduced into liposomal drug delivery¹³. These lipids have very low monomer solubility and can form, alone or in combinations with other lipids, sterically-stabilized mixed micelles¹⁴. In contrast to Amphocil micelles, which are stabilized by strong van der Waals' forces

Gene technology gems

New product developments in gene technology include 22 new probes for detecting human MER sequences and a bacterial electrode for electroporation.

LAGAN announces the availability of 22 different LaganX.PLORER probes for genomic mapping (Reader Service No. 101). These probes identify families of novel repetitive human DNA sequences called MER families. Each family is repeated 300 to 7,000 times per haploid genome. Use of LaganX.PLORER



LaganX.PLORER probes for the detection of human MER sequences.

probes through a two-step procedure can provide for a quicker contig assembly of human as well as higher primate DNAs cloned into YACs, cosmids and phages, the company says. In two hybridization steps, it is possible to go from clones to blots to overlaps. As LaganX.PLORER probes do not hybridize to rodent DNAs, they can be used in the analysis of rodent/human hybrid libraries. MER probes are supplied as double-stranded DNA at 50 ng ml⁻¹ in 10 mM Tris (pH 7.5), 1 mM EDTA.

Gibco BRL has introduced the new CONVERTIBLE Filtration Manifold System, a **clampless manifold system for the immobilization of nucleic acid or protein samples onto membranes** (Reader Service No. 102). The set-up offers a

series of interchangeable top plates and gaskets, all of which are compatible with a universal base plate. This allows the user to choose from five blot formats: two 96-well dot sizes, two 48-well slot sizes, and a 20-well slot size for processing western blots. There are no screws or clamps required for assembly of the unit. In addition, the seal between the top plate, membrane, gasket and base plate is made with the same vacuum source used to complete filtration. Gibco BRL says this results in discreet and uniform sample spots with no cross-lateral leakage of samples. The company offers pre-cut and indexed nitrocellulose and nylon membranes for use with the unit.

AMBIS, the new radioanalytical imaging system from LabLogic, is designed for high-speed, high-accuracy **imaging and quantification of radioisotopes** (¹⁴C, ³²P, ³⁵S and ¹²⁵I) on transfer membranes, gels and microtitre plates (Reader Service No. 103). The system uses a two-dimensional proportional counter that allows imaging from virtually any flat sample (up to 20 × 20 cm), including tissue sections. In contrast to autoradiography, imaging and counting are carried out automatically at high speed and with a resolution of 0.8 mm or better, the company says. When used in applications involving thin layer chromatography, the imaging of an entire TLC plate using AMBIS typically takes 20 minutes; with quantification from the CRT image taking only 10–15 minutes. This, LabLogic says, is up to 50-times faster than with conventional methods (autoradiography followed by scintillation counting). New features for TLC

within the micelle and low values of γ , these micelles are stabilized by the steric interaction of grafted polymer¹⁵, and can be, due to their increased stability¹⁵, a very attractive and versatile system for the incorporation of amphiphilic and hydrophobic drugs. □

Danilo D. Lasic is at Liposome Technology, Inc., 1050 Hamilton Court, Menlo Park, California 94025, USA. For more information, fill in reader service number 100.

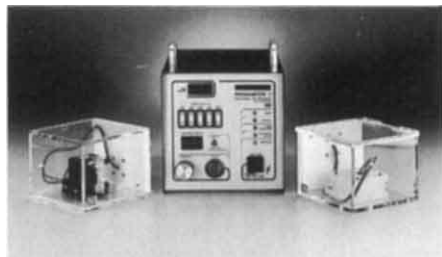
1. *Novel Drug Delivery* (eds Prescott, L.F. & Nimmo, W.S.) (Wiley, New York, 1989).
2. Brodin, A. & Nyqvist-Mayer, A. *Acta Pharm. Suec.* **19**, 267–284 (1982).
3. Guentert, T.W. et al. *Br. J. clin. Pharmacol.* **23**, 569–577 (1987).
4. Yokoyama, M. et al. *Cancer. Res.* **50**, 1693–1700

- (1990).
5. Lopez-Berestein, G. et al. *J. infect. Dis.* **151**, 704–715 (1985).
6. Janoff, A.S. et al. *Proc. Natn. Acad. Sci. U.S.A.* **85**, 6122–6125 (1988).
7. Guo, L.S.S., Fielding, R.M., Lasic, D.D., Hamilton, R.L. & Mufson, D. *Int. J. Pharm.* **75**, 45–54 (1991).
8. Lasic, D.D. *Biochim. biophys. Acta* **592**, 501–502 (1982).
9. Lasic, D.D. *Biochem. J.* **256**, 1–11 (1988).
10. Cantu, L., Corti, M., Musolino, M. & Salina, P. *Europhys. Lett.* **13**, 561–566 (1990).
11. Fielding, R.M., Smith, P.C., Wang, L.H., Porter, J. & Guo, L.S.S. *Antimicrob. Agents Chemother.* **35**, 1208–1213 (1991).
12. Fielding, R.M., Singer, A.W., Wang, L.H., Babbar, S. & Guo, L.S.S. *Antimicrob. Agents Chemother.* (in the press).
13. Papahadjopoulos, P. et al. *Proc. Natn. Acad. Sci. U.S.A.* **88**, 11460–11464 (1991).
14. Lasic, D.D., Woodlee, M.C., Martin, F.J. & Valentincic, T. *Periodicum Biologorum* **93**, 287–291 (1991).
15. Lasic, D.D., Martin, F.J., Gabizon, S.K. & Papahadjopoulos, D. *Biochim. biophys. Acta* **1070**, 187–192 (1991).

research include: a planar layout format for the assignment of origins, fronts, millimetre borders and plates (both on-screen and in printouts); auto-online, in which the user defines the background level for the automatic creation of free-form outlines around spots in the image on both one- and two-dimensional TLC plates; auto- R_F , in which the user selects R_F designation as centroid or high-count pixel of spot, or furthest point in spot from origin; and lane histograms for the quantification of percentage of lane or percentage of peaks. Other stated uses of AMBIS include: nucleic acid blots (dot and slot blots, northern, Southern and western blots), protein analysis, and CAT assays with ^{14}C .

Penetration/disruption systems

Hoefer Scientific offers a new **bacterial electrode** for use with the company's Progenitor II electroporation device (*Reader Service No. 104*). The bacterial electrode is designed to provide high transformation efficiencies of recalcitrant cells with field strengths of up to $25,000 \text{ V cm}^{-1}$. The gap between removable



Hoefer's bacterial electrode.

autoclavable electrode pins can be set to 0.2, 0.3 or 0.4 mm. The power and control unit contains a 500 V power supply and five capacitors. Continuously adjustable voltage and thirty capacitance settings between $100 \mu\text{F}$ and $2,770 \mu\text{F}$ allow the user to optimize precisely the conditions for electroporation. The discharge interval can be set from $10 \mu\text{s}$ to 9,999 ms. This degree of flexibility provides users with a choice of either complete exponentially decaying discharges or square-shaped discharges that are truncated. Any excess charge remaining on the capacitors after the pulse is delivered to the sample is discharged automatically through an internal resistor, the company says. Hoefer offers five other electrode designs for applications involving plant and animal cells.

Introduce biological particles into 'live' cells and tissues via high-velocity microprojectiles with the helium-driven **Biolistic particle delivery system** that is offered by the Genetics Systems Division of Bio-Rad (*Reader Service No. 105*). The so-called 'gene gun' packs a punch by propelling millions of microscopic gold pellets coated with DNA directly into the target cell or tissue. The helium-driven



Biolistic particle delivery system.

acceleration process used by the instrument does not require a power load or vacuum. To date, the technique has been used successfully with a wide range of intact targets, including plants, yeast, algae, organelles, cultured mammalian cells, and 'live' animal tissues *in situ*. Although the method is well established for DNA applications, it can also be used for the introduction of proteins and other molecules. Several options exist for fine tuning the system to suit specific research needs: seven microcarriers — two sizes of gold and five of tungsten — and nine rupture disks are available.

The two new **cell disruption systems** introduced by Constant Systems are designed to eliminate several drawbacks associated with present systems (*Reader Service No. 106*). Existing systems, the company says, can be limited in their output, often operating on only a single-shot basis. The two new systems, the high-pressure disrupter (HPCD-3) and the ultrahigh-pressure disrupter (UPCD-3), allow continuous or batch operation with potentially a single pass under contained, controlled conditions. Batches may require several passes to ensure adequate cell disruption and, for most cultures, the heat produced by the disruption process is harmful. The company says it has eliminated this problem by cooling the product within 0.001 seconds of it passing through the disruption head. In addition, a novel design of the disruption head eliminates the need for a high-pressure valve. This, Constant Systems says, improves reliability and eliminates the need for maintenance. The HPCD-3 unit allows disruption pressures up to 25,000 p.s.i. (1,700 bar) with a maximum delivery of 200 ml min^{-1} . The ultrahigh-pressure disrupter (UPCD-3) exhibits pressures of up to 40,000 p.s.i. (2,700 bar) and has a maximum delivery of 130 ml min^{-1} . A single 10-ml aliquot can be processed, or the systems can be continuously operated using a 500-ml reservoir. As a safeguard, the systems

incorporate an automatic shutdown facility should the reservoir empty.

Molecular miscellany

Dynal introduces a new magnetic particle-based system designed for the **colourimetric detection of amplified DNA sequences** (*Reader Service No. 107*). The system, designated DIANA (Detection of Immobilized Amplified Nucleic Acids), combines the speed and safety of magnetic particles with the specificity of a fusion protein for the identification of amplified nucleic acids, thus eliminating the need for electrophoresis, restriction mapping or hybridization assays. The target DNA may be of microbial origin, represent a genetic disorder or an allelic variation. The procedure is as follows: target DNA is amplified by a nested PCR using in-house, custom-made primers, whereby a *lac* operator sequence and biotin are incorporated; the DNA is immobilized on Dynabeads M-280 streptavidin through the biotin-streptavidin link and detected using a fusion protein that binds to the *lac* operator sequence; after the addition of a chromogenic substrate for the fusion protein, a colourimetric reaction will occur. Positive samples can be measured in either an ELISA reader or spectrophotometer, and can be used directly for solid-phase sequencing. The two-step nested primer approach minimizes the amplification of non-specific target DNA, Dynal says. Sufficient reagents are included to perform 200 qualitative or quantitative detections when performed in a microtitre tray.

Using the new pUCamp kit from International Laboratory Services, a subsidiary of Perkin-Elmer, the company says it is possible to **produce a labelled cDNA probe from a glycerol stock culture in less than four hours** (*Reader Service No. 108*). An oligonucleotide primer pair optimized for PCR allows amplification of inserts up to 3 kb long from pUC and M13 vectors, independently of the insert sequence. In this first phase, users can prepare large amounts of insert DNA from precious plasmid samples or glycerol stock cultures without the need for the tedious processes of cloning/plasmid preparation, restriction enzyme digestion and insert purification, ILS says. Once prepared, the



pUCamp kit from ILS.

amplified DNA insert can also be used as a substrate for a second phase of reactions, including the incorporation of labelled deoxynucleotides into one or both strands for use as hybridization probes, use as a template for sequencing reactions and in other enzymatic manipulations such as mapping experiments. Kit components allow reactions with one or both primers and include a pUC DNA positive control. Further kits are in development for use with other vector systems.

Rhein Biotech offers a comprehensive service covering all aspects of **yeast technology** (Reader Service No. 109). The services include: advice and consultation on culturing yeasts to obtain reliable inoculation batches, checking for contamination; examination of the identity of your yeast during the production process, and strain tagging by non-recombinant procedures. In addition, Rhein Biotech has developed the expression system of the methylotrophic yeast, *Hansenula polymorpha*, which has been successfully applied to the production of hormones (insulin, insulin growth factor), anti-clotting factors (variants of hirudin), antigens (mixed hepatitis surface antigens) and enzymes. The company will custom-tailor genetic elements for the production of a desirable protein.

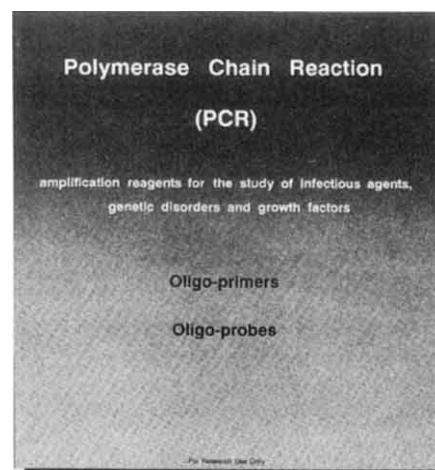
A new enzyme for megabase genomic DNA digestion — Meganuclease I — is now being offered by Boehringer (Reader Service No. 110). This 'rare cutter' cleaves

DNA within an 18 base pair recognition sequence. Because of the infrequency of the Meganuclease I cleavage site, this enzyme should be useful in genome analysis, mapping large DNA fragments and insertion mapping. Meganuclease I is qualified for cleavage of genomic DNA embedded in agarose plugs.

Products and protocols

The second edition of *ATCC Microbes and Cells at Work*, an index of special applications for microorganisms, cell lines, recombinant DNA materials and viruses offered by the American Type Culture Collection, is available in both book form and as an electronic version for use on PCs (Reader Service No. 111). This latest edition features over 1,200 new entries, 50 new pages of text and a bibliography of over 6,300 literature citations from which the applications data was compiled. The new electronic version has easy-to-use, menu-driven software for text-file searching. New application categories include: probes containing sequences for the diagnosis of disease, clones for plasmid-induced production, oncogenes, hosts for transfection/transformation, special uses for cell lines, and ATCC strains for quality control.

Genemed Biotechnologies, a company licensed by Cetus to produce polymerase chain reaction (PCR) reagents, offers its 1992 catalogue of products and procedures for PCR, Southern blotting and *in situ* hybridization (Reader Service No. 112).



Shelf stockers from Genemed.

The product lines include DNA oligonucleotide primers, DNA probes and control DNA for the study of viruses and pathogenic organisms, cytokines, growth factors, CD markers and genetic disorders. Mention is also made of Genemed's custom DNA synthesis service.

ActiClean Etox resin is a handy tool from Sterogene Bioseparations for removing endotoxins from protein or other biological solutions (Reader Service No. 113). The company says it exhibits a binding capacity of approximately 2,000 EU ml⁻¹ resin, as determined in bovine serum, which is over 25 times the capacity of immobilized polymyxin in a comparable set-up. Because of proprietary linkage chemistry, the ligand responsible for the affinity to the endotoxin is securely bound to the resin through a stable, secondary amine linkage. Consequently, leaching is less than 0.1 p.p.m., Sterogene says. In addition, the affinity of the immobilized



Sterogene's endotoxin removal system.

ligand for proteins is low, and so the mass yield of proteins from ActiClean Etox typically exceeds 95 per cent. The immobilized ligand is non-toxic, non-mutagenic and the resin can be regenerated with 1 M NaOH. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENTS

Seyvern Biotech Ltd.
CUSTOM SERVICES

- **OLIGONUCLEOTIDE SYNTHESIS**
High quality PROBES, PRIMERS, ANTISENSE OLIGO'S!
Fast-Efficient Service offering loads of modifications, BIOTIN, FLUORESCENCE, 5' & 3' labelling, Phosphothioates Affinity SupportOligos etc.
- **MULTIPLE PEPTIDE SYNTHESIS**
Epitope Mapping & Screening! We now supply 5mg quantities of high purity custom peptides up to 15 residues in length, using Fmoc chemistry. With D or L amino acids, these are supplied lyophilised in 4ml vials, cleaved from solid phase & ready for use.

Tel: 0562 825286 Fax: 0562 825284
Unit 23, Highbrook Enterprise Centre
Walsbrook Road, Kidderminster DY10 1HY

Reader Service No.17

Your Valuable Protein

MICHAEL WEST CONSULTING
Toronto, Ont.
provides top Quality Service:

- PURIFICATION • PROTEOLYSIS
- PEPTIDE SEPARATION
- SEQUENCING

CONSULTATION AND TRAINING.....ON SITE!
PHONE FAX
416-255-9288 416-979-5294

NH₂-Ala-Glu-Asn-Ser.....
YOUR SEQUENCE

Reader Service No.13

\$3.50/Base
Custom DNA Synthesis

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

KEYSTONE Laboratories, Inc.

Reader Service No. 33

Synthetic Oligonucleotides

Custom made probes and primers,
synthesised, purified and shipped in 3 days

£3.50 per base
(All inclusive)
0.2 µM scale (approx. 500 µg)

Modified bases, biotin or digoxigenin labelling,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX
Tel. XX-44-(0)71 - 326 3126
FAX XX-44-(0)71 - 733 3877

Reader Service No.56